The mind as a complex matter

Università degli Studi di Messina

Dipartimento di Scienze Cognitive

XXIX Ciclo del Corso di Dottorato in Scienze Cognitive

Dott. Andrea Zeppi

Abstract

In this thesis, we will investigate how the notion of complexity and, especially, that of computational complexity, can be applied to philosophically rich problem in order to get a better understanding of them or a straight redefinition. Differently from other more comprehensive work on complexity like that from Bruce Edmonds (1995), here we will not propose a history of said notion, or an exhaustive review. Instead, our target will be that of sticking with a notion of complexity, namely computational complexity, and proceed through comparisons and applications so that the elements of interest can emerge. This will be done with the goal of finding a philosophical role for computational complexity and to verify the hypothesis that this particular notion of complexity is particularly well suited to evaluate the plausibility of all kind of theories.

The thesis will be ideally divided in three separate sections. In the first we will analyse the philosophical aspects of complexity. We will see what kind of features does it have as a notion and why is it important for philosophy of mind and cognitive science. In the second section, we will look at three different philosophical applications of computational complexity and of the tractable cognition thesis. The third section will be dedicated to more cognitivistic application. The rationale behind this section is to look at how computational complexity improves the understanding of cognitive capacities and features of cognitive systems that have high philosophical relevance. This allows to both see how the role of plausibility notion is fulfilled on the part of computational complexity both on the upper bound and lower bound of cognition.

Table of contents

Introduction

Philosophy of mind has been throughout its (not so long) history has been obsessed with theoretical gaps. One example for all comes from the most classic of philosophical debate: the mind-body problem. This problem rapidly gave rise to its peculiar gap and pretty famous gap, the explanatory gap (Levine 1983). From there on we had seen the rise of hard and easy problems, some reduction, a few elimination attempts and even a supervenience. However, despite the pretty fertile ground for good philosophy, this is not a thesis about gaps. This is a thesis about notions that sprout around gaps and, in particular, this is a theory about complexity. However, differently from other more comprehensive work on complexity like that from Bruce Edmonds (1995), here we will not propose a history of the evolution of said notion, or an exhaustive review of the contemporary approaches in the various philosophical and scientific fields. Instead of looking at all the possible meanings of complexity (rigorous and vague), our target will be that of sticking with a notion of complexity, namely computational complexity, and proceed through comparisons and applications. This will be more of a philosophical tinkering than an analysis, but it will allow us to clarify how a specific notion of complexity fits inside the theoretical landscape of cognitive science.

The thesis will be ideally divided in three separate sections. In the first we will analyse the philosophical aspects of complexity. We will see what kind of features does it have as a notion and why is it important for philosophy of mind and cognitive science. This first part will ideally represent our theoretical yard, where the foundations of our next philosophical task will be set. A space in which we will present the basic components of our

philosophical framework and has the main scope of grounding a specific formal notion of complexity, namely computational complexity as a theoretical "ruler" for measuring the plausibility of theories, especially that of cognitivistic theories. In the second section, we will look at three different philosophical applications of computational complexity and of the tractable cognition thesis. Since our analysis will strongly rely on comparison, evaluating notions well known to philosophy may reveal a lot and provide with more solid elements to understand how computational complexity behaves as a philosophical notion and how it scales accordingly. The third section will be dedicated to more cognitivistic application. The rationale behind this section is to look at how computational complexity improves the understanding of cognitive capacities and features of cognitive systems that have high philosophical relevance. This allows to both see how the role of plausibility notion is fulfilled on the part of computational complexity both on the upper bound and lower bound of cognition.

The first section will articulate in two chapters. In the first we will explain why have we chosen complexity as the main topic. This problem will be tackled by proceeding incrementally. We will first argue in favour of the philosophical relevance of complexity and then look at why Computational complexity should be considered relevant not only to the scientific study of cognition, but also to the philosophical side of the cognitivistic program.

The second section will comprise of the third, fourth and fifth chapters. In the third we will take a step into the key element of complexity by considering the important relationship that it entertains with simplicity. The choice of simplicity comes from three main reasons: the notion of simplicity has a strong link with that of complexity, long philosophical history and has found applications in cognitive science. In the fourth chapter, we will instead look at how two type complexities can be devised behind two well-known approach to the explanation of cognition: dynamicism and computationalism.

The aim is to see what different notions of complexity reveals about the theories that make use of them and, also, how computational complexity has improved the computationalist framework. In the fifth and last chapter of this section we will confront computational complexity with Tononi's formal notion of complexity. A notion that has found both a cognitive and philosophical application.

The third and last section will consider two more cognitively flavoured applications of computational complexity. First we will see how the tractable cognition thesis can be applied to the mindreading cognitive capacity in order to search for new and more cognitively plausible version of it. In the second part, we will instead look at a different way in which computational complexity can be applied to cognitive science. In particular, we will try to indicate a notion of "minimal complexity" that accounts for the very low level characteristics that a concrete computational system must possess to sustain cognition.

Each chapter will open with a short introduction to the topic that they cover and a brief summary on how that part fits in the bigger scheme of the thesis. Inside every chapter, section and subsections will be titled and numbered, in order to ease the work of the reader.

Acknowledgments

Before the real thesis starts. I would like to acknowledge my supervisors, who have advised me during the last three years, my parents Ivano, Letizia and my brother Duccio, who have supported me for even longer, and my girlfriend Benedetta who has always been at my side despite the physical distance that often separated us. A last word of gratitude goes to those colleagues and friends that contributed in some way to the writing of this thesis both through suggestions and support in time of need.

Notes to the reader:

Chapter 3 is inspired by "A.Zeppi (2015) Dalla Complessità alla Semplicità, Conference: XXIV Convegno nazIonale deI dottorati di ricerca in filosofia".
Chapter 4 is inspired by "A. Zeppi (2016) Two ways into complexity, in Language in Complexity - The emergence of meaning, Edition: 1, Chapter: 9, Publisher: Springer, Editors: Francesco La Mantia, Ignazio Licata, Pietro Perconti, pp.148-158.
Chapter 6 is inspired by: A.Zeppi (2014) Defining Tractable Mindreading, conference paper: AISC-CODISCO 2014 Roma3, At http://issuu.com/neascience/docs/atti_aisc_2014/1, Volume: NeaScience Anno1 Vol.5- Atti Aisc 2014; P.Perconti, A.Zeppi (2015). Mindreading and Computational Tractability. Anthropology & Philosophy, 11 - 2014/2015, ISBN - 978-88-5752-826-7;
Chapter 7 is inspired by: A.Zeppi, A.Plebe, P.Perconti (2016) Looking at complexity the other way around, conference paper: MIUCS 2016 Warsaw.

Chapter 1

Why should philosophy care about complexity

In this chapter, we will tackle the specific issue of the relevance of complexity for philosophy and introduce the special case represented by philosophy of mind. Complexity is a term that, funnily, is complex in itself. It refers to a riddle of concepts and notions that, when considered separately, may indeed seem inconsistent. In order to avoid this discouraging fact, we will, in the following paragraphs, focus on evaluating complexity through its applications. In light of this principle this chapter will articulate in two parts. The first one will be focused on how complexity behaves as an informal notion There we will look into the various peculiarities of complexity by looking at a number of philosophical applications of it. The list examples provided will not be exhaustive, since it is not the aim of the present thesis to give a general account of complexity. However, by proposing a number of relevant cases we will assess the specific philosophical relevance of complexity and justify then why a better theoretical understanding of complexity should be pursued and what requisites should satisfy. In the second part of this chapter we will instead concentrate on the formal measures of complexity that are been proposed until now in the literature. Again, the list will not be exhaustive, but instead comprehensive and the focus will be on the applications of complexity more than on the possibility of providing a unified definition of it.

In the remainder of the chapter we will draw our conclusions by adding up the considerations made in the single sections. In particular, we will suggest that the only way of fruitfully apply complexity is to accept its

multifaceted nature and to choose a specific application of it. Where for application we not only intend a field, but also a theoretical role inside that field. The specific philosophical relevance of complexity, we argue, stands in the in the process of reaching sufficiently clear and rigorous notion of complexity that can be fruitfully applied to a particular field, in this case cognitive science and philosophy of mind, so that an improvement of the accuracy and explanatory quality of scientific and philosophical theories can be obtained.

1.1 Introducing complexity

Complexity is often described as an ambiguous and vague notion, or at least a difficult one to approach systematically. The reason behind this judgement stands in the fact that this notion appears in wide range of arguments and thesis and, as Morin observes, this may also be the case why "*the expression "it is complex" in fact expresses the difficulty of giving a definition or explanation*"(Morin 2007, p. 2). However, the obscurity for the notion also has a slightly different meaning. Complexity takes not only the shape of an informal criterion, but is also the centre of numerous formal and rigorous "measures" of complexity that have found usage in a spectrum of technical and scientific fields. Such technical notions of complexity are often obscure in their own right because they are also difficult to approach and understand. Also, the linguistic use of the terms "complexity" and "complex", scattered over a number of different applications and fields, clearly indicate the fact that the notion which they refer to is not unique or coherent. So, a first preliminary answer to the relevance question is that a philosophical analysis of complexity may be useful for clarifying the interplays between wildly different informal an informal version of the notion. This qualifies complexity as a genuine philosophical topic indeed, but doesn't say much for the

relevancy of it in the panorama of contemporary cognitive science. The implicit meanings suggested in informal complexities are important for reaching the right balance between the constraints of formality and the richness of the phenomenon that we aim to explain. Also, the theoretical nooks and crannies, as we will see, of the "formal vs. informal" comparison call for theoretical expertise of philosophy. Relevancy however may come from the fact that clarifying complexity, while at the same time catching some of its less rigorous nuances, may indeed be of profit in the actual process of doing science. So, since complexity appears in a lot of philosophy, but also in a lot of science, clarifying and testing the various forms of it in order to find not the best one in general, but the best one given the application would be a nice contribution.

In order to take into consideration these issues we will proceed as follows: first we will provide a short review of the various informal approaches to complexity that have already been proposed in the literature. Such an approach differentiates us from other more exhaustive reviews of the formal notion of complexity, such as that of Bruce Edmonds (1995), and will provide a number of insights that, we argue, all together constitute the implicit "message" of complexity. Then, our second target will be that of formal notions of complexity. There we will present different alternatives that have been proposed in the literature. We will see why they are called "measures" of complexity and what are the key notions that have been proposed as philosophical counterparts of them. Here another way in which we should justify the relevance of complexity emerges. Formal measures of complexity, like we said before, are not less obscure than the informal one, but their arcane nature comes from the sheer difficulty of approaching the technical know-how needed for handling the notions themselves. To justify that these solutions a turnout is needed and, as we will see, clarity and rigour is not always enough.

In the end, formal and informal notions both have a purpose, especially a philosophical one, and the multiform and widespread use of complexity is indeed a sign of that. A clear and rigorous understanding on how complexity fits in contemporary scientific theories needs such a fine analysis, especially if complexity is a notion that keeps to be used in cognitive and other sciences in many different ways.

1.2 Informal complexities

Since our main target is indeed philosophy of mind we will start our short review by looking at a emblematic topic that made extensive use of complexity. Complexity takes an unquestionable important role in some famous arguments against the reduction of consciousness (Dreyfus 1972; Chalmers 1996; Dreyfus 1992). Here complexity plays an important role and it is often used by the supporters of non-reductionism to underline how the various attempts at naturalizing consciousness fails to catch the "richness" of first-person experience. What we may find is an implicit and allusive use of the notion, but when it does emerge fully and explicitly emerge an interesting pattern shows up. This is the case, for example, of John Searle's argument against computationalism:

> *"For any sufficiently complex physical object O [. . .] and for any arbitrary program P , there exists an isomorphic mapping M from some subset S of the physical states of O to the formal structure of P. " (Searle 1990)*

Here complexity is used as cornerstone for underlining the weakness of the computational theory of mind, but it is the use of "sufficiently" in conjunction with "complex" that should draw the attention. This particular style of arguments provides us with two additional aspects of complexity: the first

one is that being complex is a prerogative of an entity of some sort[1], the second element is that such feature seems to work as a requirement for the theory that underlies the possibility of accepting it or not. This requirement can be interpreted as global propriety that an explanation, or a theory, should address by principle a that, at least for the moment, we can trace back to a loose notion of "richness". If said explanation is not rich, or complex, enough it should be considered as unsatisfactory and then removed or at least rephrased in other, more appropriate terms.

A similar approach to the one above can be found in the dynamical hypothesis to the explanation of cognition (Malik 2002; T van Gelder 1998; Horgan and Tienson 1992). In this account complexity is used to highlight the fact that the actual neuronal and cognitive dynamics cannot be satisfactorily caught by the classical computationalist approach and therefore this last approach doesn't represent a theoretical acceptable framework. This particular aspect of the dynamical hypothesis clearly emerges from the following statement by van Gelder:

> *"The claim is that we must understand cognitive agents as dynamical systems, because only in that way will our account of what cognition is be properly integrated with our account of how the world sustains any of it."(T van Gelder 1998, p. 623)*

Since the focus is on dynamical systems that are complex in virtue of a set of characteristics here complexity has to be again interpreted as an inherent property of the actual object. However, dynanicism makes a crucial step forward for the sake of our analysis and by taking a closer look at complex systems we can see that their essential characteristics sheds some light on the numerous aspects of complexity that before were only loosely expressed.

[1] In case of cognitive science a cognitive system.

These systems are in fact so because they express a number of properties: like that of self-organization and emergence (Dixon et al. 2012; Gibbs and Van Orden 2012; Yamashita and Tani 2008; Funtowicz and Ravetz 1994), being concrete not idealized systems (T van Gelder 1998), being interactive and mostly holistic in their conception. These characteristics indeed provides us with several examples of what complexity may actually mean and because they are indeed related to the "richness" aspect that we already underlined. Complexity comes from interactivity, being immersed in a continuum and, crucially, from being actually concrete and immersed in a context. Here we will not go further in the analysis since the different conceptions of complexity that are employed in computational and the dynamical frameworks will be the argument of chapter three, but suffice to say that dynamicism addresses in a more detailed way the fears of the supporters of non-reductivism by injecting richness in cognitive science and, then, addressing the system notion itself.

These aspects of complex systems that ground dynamicism undoubtedly resonate with a philosophical proposal entirely dedicated to complexity, that of Edgar Morin. Morin's complexity, is informal and used as cornerstone for its theoretical framework. Here complexity is still opposed the reductionism of scientific practice, too focused on forcing the richness and variety of phenomena into explanations that are inherently disjointed and simplistic[2]. This comes ad evident if we consider the following quote:

> *"The problem is not to create a general theory covering everything from atoms, molecules, and stars to cells, organisms, artifacts, and society. Rather, the problem is to consider atoms, stars, cells, artifacts, and*

[2] Here the literature on Morin's paradigm of complexity often refers to "simple", here we preferred to use the word "simplistic" in light of the relation between the two notions of complexity and simplicity that we will explore in chapter 2. Furthermore, as we will see, simplicity itself has a deep history of philosophical development that doesn't agree the negative connotation usually given in Morin.

society-that is to say, all aspects of reality, including, and in particular, our own – in richer way in the light of the complexity of systems and organization" (Morin 1992, pp. 382-383)

In Morin's framework complexity is then an intimate property of the real world and should then be accounted for in the scientific practice. Even more important is however to observe that the complexity of phenomena calls also for a revision of the theory. Explanations, hypothesis and other theoretical elements that ground the practice of building scientific knowledge are also a crucial part of the equation, so complexity has to become not only an important feature of systems, but also a critical feature of theories. That because building meaningful scientific knowledge is intimately grounded on the process of reaching the crucial balance between the constraints of science and the richness of phenomena.

If preliminary conclusions can be drawn from this short list of examples is that the informal notions of complexity here considered all point to the following question: what are the necessary and sufficient properties that make a system complex enough to be considered a cognitive being?

The path that we followed started with Searle and that example clearly set the ground for individuating the question itself. Then, by taking dynamicism into the equation we have seen an evolution of the concept of complexity as a feature of systems. There the various components of complexity were not implicit but could be explicated in the series of features that a system needs to possess in order for it to be considered complex. In the end, we added, with Morin's complex thinking, a new crucial dimension to the problem. Through this addition becomes evident that not only the complexity of systems should be considered, but also that of the theories that try to explain them need to be.

However, a common element between these two way of conceiving complexity exists: complexity, regardless of its type, is always a global

property about the fact that a system, or a theory, express a set of necessary features that make such system, or theory, acceptable. This particular property makes possible to recognize in complexity a theoretical behaviour that can be associated to that of a principle of plausibility. A principle that in relation of other topics has been defined as follow:

> *"the degree of probability that a model is accurate in the existence of, and distinctions between, the various entities and activities it postulates" (Gervais and Weber 2012, p.* 140*)*

Even if this particular definition applies to models, the same rationale can be translated for systems and theories so that complexity can be effectively conceived as performing the role of a plausibility principle. The complexity of systems can be intended as a feature that, when accounted for, also improves the accuracy of the explanations that are tied to them. On the other hand, this logic also applies to theories and can be used we maximise the precision of predictions and, ultimately, the quality of the explanation.

How to consider critically all these elements sure is philosophy, and this may actually be one of the answer to the question in the title, but we should not be satisfied. One of the main targets for reaching a better understanding of any phenomena, and cognitive ones nonetheless, would be that of reaching a clearer understanding of the role of complexity by making complexity itself a more rigorous notion in the first place. One of the way in which this can be done is through a transition from informal, and somewhat ambiguous, definitions to formal ones. That because, while we recognize that the informal kind has indeed the role of pointing to the requirements that theories have to express. If we continue to apply a vague notion of complexity, we would also run the risk of missing some crucial interplays between complexity and some philosophically rich and useful notion. Furthermore, the special case of the transition between exclusive philosophical speculation and interdisciplinary

cognitive science, definitely calls for more rigorous and coherent way of defining complexity. In the following paragraph, we will address therefore the formal notions of complexity and see how they behave in comparison to the informal ones that we have just seen.

1.3 Formal complexities

Differently from the argument of previous paragraph the topic at hand surely needs a different approach. Technical and formal notions of complexity, as already said in the opening of the chapter, are not obscure because they lack precision, but because they need to be approached with a technically keen eye. One element of our approach will not, however, change. We are still not interested in a full review, so in here we will again proceed by mentioning some relevant and see if and what are the philosophically rich topics to which they relate to. A preliminary remark is however in order. The formality of the notions that we are going to consider is expressed though the metric element of these complexities. The fact that a theoretical element can be used as a measure of some kind of feature is indeed a big indication of its rigor and precision. That usually come at the expense of the generality of the principle, else said in the possibility of applying it to different problems and phenomena. However, this specialization factor is balanced by the sheer number of the various measures that have been proposed in history, so that for a given application the right measure of complexity can be used. The metric use of formal complexities is not only a taxonomic nuisance however. It indicates a general trend that we were able to observe also in informally defined complexities: that of using complexity to express globally some local properties of a target entity in order to make a comparison possible. Before making more claims it should be useful to look at the various measure that

have been proposed not only in cognitive science, but also in other sciences as well.

A useful guide into the various type of measures of complexity is provided by Seth Lloyd of the MIT (Lloyd, n.d.). In his non-exhaustive, but pretty comprehensive list, he provides a synthetic view on the various measures of complexity available, but also a taxonomy that is indeed useful in framing the issue at hand. Lloyd recognizes that the measures of complexity can be, and has been also previously (Edmonds 1995; Löfgren 1973; Löfgren 1977), categorized in the following way: difficulty of description, difficulty of creation and degree of organizations

The first category mentioned in the list is that of difficulty of description. Shannon's information (Shannon 1948) and Kolmogorov complexity (Solomonoff 1964; Kolmogorov 1965; Chaitin 1966) are the most important among the listed measures of complexity and the rationale under this categorization becomes evident when their nature is considered. On one side, we have Shannon's measure of information, or also called Shannon's entropy. This notion is a measure of the quantity of information that's transmitted between a source and a receiver through a channel. The measure make use of statistics and probability to formally define an intuition that is well explained by the following quote:

> *"The intuitive idea behind Shannon's measure is that the more surprising a message is, the more information it conveys. If I tell you that the sun will rise tomorrow, this is very unsurprising. But if I say that it won't, this is very surprising indeed, and in some intuitive sense more informative."* (Dunn 2008, p. 590)

In SI sense information is then directly proportional of surprise and then the inverse of the probability that a certain event will occur. Through its logarithmic definition Shannon's information realizes an encoding from the

inverse of the probability to strings of bits. It comes easily at this point the link between this measure of information and its categorization as a measure of descriptive complexity. The more an event is complex, the more information will be needed for determining its outcome and then the lengthier the description will be. On the other side, we have Kolmogorov complexity (KC), also called Kolmogorov-Chaitin complexity. This type of complexity is also a measure of information, but based on the length of the description, or an encoding, of a given object[3]. The fact that it is also listed a form of descriptive complexity should not come as a surprise However, by looking at some details of it we can better understand why Lloyd adds the element of difficulty to its way of categorising complexity in general.

KC varies in function of "the size of the shortest program that, without additional data, computes the strings and terminate" (Vitanyi 1998, p. 2). This way of seeing information is based on the intuition that the structure of a finite string (often thought as a binary string) actually expresses a precise characteristic when such string represents an algorithm that is computed by a Turing machine. If a string shows systematic regularities than it also means that it can be restated in a shorter way without altering the effectivity of the computation. Such string will contain a small amount of information if the redundancies in it high and a large amount of information if, instead, it shows a less predictable structure. That happens regardless of the actual length. So, what KC provides is a measure of the actual structure of a string through the evaluation of the difficulty of making that string shorter. This is also one of the reasons why KC has close ties with applications in computer science like data compression (Li and Vitányi 2009) and, more importantly for us, also with topics in philosophy like simplicity (Rissanen 1978) where the notions of size and minimum description size (Edmonds 1995) play an important role.

[3] Usually a string of symbols, a program or an algorithm.

This particular relationship between KC and the epistemological notion of simplicity also underlies the general relation between complexity and simplicity. In fact, these elements resonate also in cognitive science and this is the reason why we will dedicate the chapter 2 to investigate them.

The second category that Lloyd recognize is that of complexity as difficulty of creation. In this category, the focus shifts from the aspects of the length, size or structure of descriptions to the actual effort required to perform a certain process or task. This is also why this type of complexity often are measured in resources like energy, costs and computational resources. Among the different measures listed under this class, computational complexity is indeed one of the most iconic and widely adopted. Computational complexity (CC) is a measure of the hardness of computable functions, it offers a measure of the resources that are needed to solve a computational problem, where a computational problem is indeed expressed as a computable function. For this reason, the focus of CC is not on what is computable in principle like for computability theory, but on what can be computed in practice and can then be considered computationally tractable. The resources that are considered by this measure of complexity are time and space (memory in particular). If a problem, or a function, takes an unreasonable amount of time (or space) to be solved it is considered intractable, or non-computable in practice, the contrary happens if instead the resources needed remain under a certain threshold. Since this type of complexity will be the main topic of the next paragraph here we will not provide further technical details. However even the few elements that we have collected are enough for individuating the purpose of CC into the philosophic framework that we are exploring. The link with difficulty comes as particularly evident since CC is indeed a measure of computational effort. One advantage of CC is that it can be easily applied to cognitive science, if a computational explanation of it is accepted, and provide a link between the

specific notions of computational tractability and that of cognitive feasibility. Another important philosophical notion that can be linked to CC is that of parsimony, an aspect of simplicity. This particular application will be investigated in details in the second chapter, so for the moment we can go on and look at the third and last category, that of the degree of organization.

Under such class we find complexities like: sophistication, conditional information, channel capacity and mutual information. These are only a few of the examples that Lloyd gives in his list and at least the abovementioned one are special case or further refinement of SI or KC. If we consider the case of channel capacity, we can clearly see on example of that. Channel capacity can be defined as "the maximum number of distinguishable signals for *n* uses of a communication channel" (Cover and Thomas 2006, p. 184) is then equal to the maximum mutual information between a source A and a receiver B. Where for mutual information we intend the measure of the amount of information of one random variable that is dependent on some other random variable. Mutual information measures then how much the knowledge about an event modifies the outcome of another event, where every event can be codified using SI. Now, if we move from the description of the measures themselves and focus on the features that they express we can see that such measures are indeed well suited for expressing notions like, order, disorder, organization articulation and integration. This makes them particularly suited to for the evaluation of the complexity in systems, architectures and structures. One application in particular, Tononi complexity, is particularly interesting for us, because it provides a measure of complexity that is aimed at the human brain and tailored toward the characterization of the necessary features for expressing human-level consciousness. Given that we will take a detailed look at it in chapter 4.

Having considered the scenery provided by formal measures of complexity we can indeed draw some preliminary conclusions before we

wrap up the section. We started the paragraph by saying that formal complexities have the property of being rigorous. The rigorousness is indeed balanced by the fact that they are often special-purpose and tailored with a particular application in mind. However, as we have seen this hasn't impaired the various notions to also have philosophical application. So, if you set aside the ambition of finding a unified notion of complexity, you only have to choose the right notion that applies to the particular target framework. In line with this predicament, in the next and final paragraph we will take the peculiarities of cognitive science into the equation. In order to do so we will first provide an answer to the question about the philosophical relevance of complexity and see how, by using this answer, it is possible to approach a notion of complexity that can apply to the particular case of cognitive science.

1.4 Wrapping up

At the beginning of the present chapter we present the hurdles posed by notion complexity, especially its vagueness, and introduced how it should be reconceived under the light of the relevance problem. We then considered the informal and formal types of complexities. From the analysis of the first type we concluded that complexity can be predicated both to systems and theories. Also, we have seen that this kind of distinction happens to be relevant in arguments revolving around the plausibility of theories. On formal side of complexity, we have seen how the different measures solve the original vagueness of informal depictions of complexity while at the same time losing in domain-generality. More precise measures of complexity need then also more refined framework for them to be applied correctly and fruitfully.

All these elements stack up into an answer to the original question on why should philosophy care about complexity. Some preliminary reasons to

believe that complexity is indeed of philosophical significance come from the wide philosophical use of complexity, but also from the fragmentation of it. This partial proof of relevance follows from the sheer number of topics and fields that complexity touches, but what also indicates is huge need for theoretical clarification. However, the meaning of "relevance" can be interpreted in two ways and, while the difference between the two is subtle, we will propose the following way of stating it: On the one hand "relevance" can be intended as being able to qualify for philosophical analysis as a whole; On the other hand, we have instead relevance intended as opportunity of philosophical research. The little gap between these two versions of the same concept stands in the fact that even if a problem or a topic has indeed the characteristics of being philosophically interesting as a whole, it may also be the case that in the philosophical subfield (like philosophy of mind) that problem may be irrelevant or at least not theoretically profitable. This is of course not the case and again a preliminary confirmation comes from the fact that among the example that we have proposed the majority indeed comes from the scientific or philosophical study of cognition. The difference between informal and informal notion of complexity is another gap that indeed calls for a philosophical exploration. While informal complexities seem to be applied in argument that polemically individuate the explanatory requirement for complexity, the formal measures of it provide the tool to reach a clearer understanding of the notion to which they are applied. How to reconcile these two apparently different perspective has been and still is a philosophical endeavour. Bruce Edmonds (1995) for example adopts a general definition of complexity[4] that ties this notion to language and, ultimately, to the difficulty of reaching definitions and formulations for certain phenomena. Differently from this type of approach we support the following alternative:

philosophy should address the gap between informal and informal notions of complexity by exploring how the latter can be fruitfully applied to known philosophical notions and problems. By putting complexity at work, we will sure have hints at how it performs as a philosophical problem in itself, as well as have the opportunity of having a new perspective on different problems. As we will see in detail in the following chapter, the philosophical research on complexity should be considered valuable for reaching a better understanding and the relationship between concept like plausibility and richness (Gervais and Weber 2012). Also, distinguishing between the complexity of systems and the complexity of theories is indeed of crucial importance, especially in a field like cognitive science that have systems as their target. Cognive science studies cognition and those systems that performs it. In this regard the complexity notion has to be evaluated with this particular application in mind. The distinction between the complexity of systems and that of theories is one of the possible taxonomy of complexity that applies pretty well to the cognitivistic framework. Accounting for the complexity of cognitive systems is one of the requisite that imposed to plausible theories of cognition by theoretical frameworks like the dynamicist one. In the same regard, then, the complexity of the theories that explain cognitive phenomena should be taken in high philosophical regard, especially if the possibility of a formal definition of it is possible. In the next chapter, we will move onwards and apply the methodology that we have just drafted. A specific formal notion of complexity will be chosen and its relevance will be evaluated.

Chapter 2

Why should philosophy of mind care for computational complexity?

Previously we considered how complexity, as a whole, has indeed all can be considered not only a genuine philosophical problem, but also a relevant one nonetheless. In the present chapter, we will instead unravel our cards and endorse a specific formal notion of complexity, *viz*. computational complexity. This permit us to evaluate in depth a single notion, present it, see how it has been applied in cognitive science and look at what role does it fulfil in it. In order to do that, we will use the article by Aaronson "*Why philosophers should care about computational complexity*" (2011) as a starting point for considering the special case of cognitive science and suggest an update that tackles the specific problems that the explanation of cognitive phenomena presents. This will also concede us the opportunity of presenting the "Tractable Cognition Thesis" [5] (van Rooij 2008; Tsotos 1990; Frixione 2001). The goal in this chapter will be to argue that the relevance of computational complexity for both cognitive science and philosophy of mind stands in the fact that it can be used as a "plausibility" notion. A notion that, when applied in a sufficiently clear and rigorous form makes possible to improve the accuracy and explanatory quality of scientific and philosophical theories. It does that by accounting for the inherent limitation of cognitive systems, on the one hand, and introducing a threshold for discerning between cognitively plausible and implausible tasks, on the other hand.

[5] TCT from now on.

2.1 Introducing computational complexity

We already mentioned how complexity can be obscure in at least two senses: one because of the eventual vagueness of its definition, the other one for the sheer technical difficulty of approaching the subject. but also from the conceptual hurdles that this complexity theory hides. That also explains the need felt from Aaronson (2011) to explain "Why should philosopher care for computational complexity". However, even if the examples provided by Aaronson are indeed comprehensive, we think that something is amiss. What we want to argue is that computational complexity has found one of the most prolific field of application in cognitive science, especially when sided with theoretical considerations about the nature of computational explanation of cognition. However, before discussing examples that Aaronson gives and provide our own considerations and updates, we should provide the reader with at least an essential introduction to computational complexity.

The core of computational complexity stands on the distinction between effective computation and efficient computation. While effective computation indicates those functions that can be computed in principle, falling then in the spectrum of computability, computational complexity applies to functions that are computable in practice. In terms of philosophical categories computability deals with the actual possibility or impossibility of a certain mathematical function. If the problem, and the function that models this problem, can be computed by a Turing Machine then it also means that it is computationally possible, otherwise if it's not. Computational complexity instead, by dealing with those functions (or computational problem) that are computable in practice, is interested in the contingencies of computation and then in those constraints that make a certain problem computationally efficient. and in the way in which what takes the name of computational tractability can be categorized and explored. Even if the computational

framework is now widely accepted in cognitive science, further clarifications have to be done in order to define the differences that separate standard computationalist solutions interested in effective computations from those that, taking into considerations tractability issues, are instead committed to efficient cognitive computation.

First of all, what tout-court computationalism claims (in cognitive science) is that cognitive or mental processes are computations (Cordeschi and Frixione 2007; Piccinini 2009), hence that a cognitive process always expresses one or more computable function[6] (Massaro and Cowan 1993; Anderson 1990; Cummins 2000), that is, functions that are computable by the means of an effective procedure (an algorithm). However, even if it's well known that not every function is Turing-computable and that there are more or less precise limits to what we can call computational, that does not seem to pose too much of a boundary. Computability checks are, for instance, inherently unbounded in a number of ways, and that is because they rely on an idealized computational model (the Turing Machine) that, *de* facto, can possibly rely on infinite time and space resources (an infinite tape and infinite computational step). A function is computable in the classical sense if exists a finite effective procedure, hence an algorithm, that halts after a finite number of steps, uses a finite number amount of storage and works for arbitrarily large set of inputs (Enderton 1977). This kind of solution has to be somewhat finite, but there's no *a priori* indication about how much memory or time does it needs in order to return an output. A classic example of non-computable function is the halting problem (Turing 1936), a decision problem according to a Turing machine having as input a program should output if said program

[6] Here the received view (classic computationalism) mainly indicates Turing-computable functions as a reference (Marr 1982; J. Fodor 1975), but other proposals tried to suggest that cognitive processes (and especially those present in the human cognitive system), even if computational, are not to Turing-computable and then must rely on some other, maybe more powerful (Wegner and Goldin 2003; Copeland 2002; Steinhart 2002), type of computation (Tim Van Gelder 1995a; Penrose 1999).

will correctly halt (and actually output something) or instead continue to run forever.

Computational complexity ideally takes a step further and continue the analysis from where computability theory has left it. As we said earlier the propriety in which computational complexity is interested is not computability tout-court, but computational tractability. This propriety, tractability, is possessed by Turing-computable functions that can be computed using a certain amount of computational resources. The computational resources here considered are the following:

TIME: the time in term of computational steps that a function will require in order to return an output;
SPACE: the amount of memory (the length of the relative Turing machine tape) that a function will require in order to return an output.

However, these two computational resources are not evaluated directly, like happens for example in information theory, but in function of the size of the input[9]. The size of the input $|i|$ can be defined as the number of symbols on the tape of the Turing machine that is used to represent the computational problem and then the function. Another proof of the indirect nature of the measurement obtained through computational complexity stands in the asymptotic notation $O(f(x))$, called in this case *big-Oh notation*[10], that is used to classify computational problems according their rate of growth in function of the input size $|i|$. The classification on which the tractability threshold is base can be exemplified as the following: if we assume a computable function M allora, if for every possible input i M is decided in a number of steps

[9] One of the requirements that are imposed to the input is that of having a reasonable, non-redundant enconding. For an analysis of the issues behind this assumption look Kwisthout (2012).
[10]

$O\big(t(|i|)\big)$, then we can say that M is decided in $O\big(t(|i|)\big)$ time and has a T-complexity (time complexity) of $O\big(t(|i|)\big)$. This type of notation has also the practical purpose, being it a measure of the worst case scenario (where all the possible input are taken into the account), to allow the overlooking of differences in running time coming from constants and lower degree polynomials (Van Rooij 2008; Frixione 2001; Arora and Barak 2009). This insensibility to certain negligible elements of functions running time is also stated by the following thesis:

Invariance thesis: given a "reasonable encoding" of the input and two equally reasonable Turing machines M_1 e M_2, the complexity of a certain problem ψ_T for M_1 e M_2 will only vary of a polynomial (Garey and Johnson 1979).

What the thesis of invariance allows, then, is to talk about the difficulty of problems without having to take into the equation also the performance of the actual computing system or model that solves them. Now that we have the thesis of invariance and the specific asymptotic notation we have all the tools needed for categorizing computational problems according to their inherent difficulty. Computational complexity subdivides computable functions in a taxonomy of complexity classes based on the two computational resources of TIME and SPACE. Here we will only consider time complexity. That not only for space reason, but also because time and space complexities are intimately related. Time complexity already expresses a certain measure of space complexity. (Garey and Johnson 1979; Van Rooij 2008). On fundamental distinction is that between functions that can be solved in polynomial PTIME (for example $O(t(|i|^{\alpha}))$ where α is a constant) and those that require instead exponential time $\text{EXPTIME} = O\left(t\left(\alpha^{|i|}\right)\right)$.

Traditionally polynomial time P is considered as the reference threshold for efficient computation. A problem that allows for a worst-case solution in polynomial time is considered then computationally tractable (Garey and Johnson 1979). Vice-versa is classically understood as intractable a computable problem that does not run in polynomial time. The distinction between tractable problems (also called "easy to solve") and non-tractable problems does not represent the entire spectrum of the taxonomy of classes that populate computational complexity. The kind of problems that we have considered those that can be decided in deterministically in PTIME. This means that a deterministic Turing machine (DTM) will decide their output without any branching, every state will determine univocally the transition to the following one. However, when we shift to a non-deterministic model of computation, like non-deterministic Turing machine (NTM), a new complexity class of problems appears: polynomial non-deterministic time NPTIME. This complexity class is also called of the problems that are hard to solve but easy to verify. That because under this computational model the transition between one state to another is no more deterministic and then univocally defined. Branching can occur and this allows for choices that don't always lead to tractable running times. If a solution to such problems is available, then it will be easy to check if this is actually the case. Furthermore, since a DTM is a special case of NTM we can conclude that the set P of problems solvable in polynomial time is included in NP ($P \subseteq NP$)[17]. The topography of complexity is however still incomplete. Two intersections between the P, NP and EXP classes have been individuated: that of NP-hard

[17] While this is now considered a triviality, the equivalence between these two classes of problems is an open question, and is in fact one of the most important contemporary challenges of theoretical computer science with a one million dollar price for whoever comes out with a solution (Cook 2003). The reason for this is easy to illustrate: If P=NP that would mean that a polynomial solution is available for every algorithm or problem that is now believed to be in NP, with heavy ripercussions for computing in general.

problems and that of NP-complete problems. The first of these two classes individuates those problems that have a difficulty that is "at least equal to that of every other problem in NP"[18]. The second in instead relative to the class of most difficult problems that belongs to NP, and then to those problems that are both NP-hard and NP. The present distinction, along with the ones that we have previously illustrated, allows to enhance the classifying capacity of computational complexity and its applications. One last remark has to be done: while computational complexity fits snuggly inside Marr's levels of analysis (Marr 1982), since it is a theory that targets computational problem its applicability is almost always limited to the first level, the computational one.

In the next paragraph we will consider one the possible application of computational complexity, and see how the taxonomy that it provides can be used to apply a plausibility threshold to theories and models of cognitive capacities.

2.2 Going further: The tractable cognition thesis

At the start of the previous paragraph we mentioned Scott Aaronson and his attempt to attract the (philosophical) attention on computational complexity. In his paper, he goes through various example ranging from the relevance of polynomial time for philosophy of computation, the applications of computational complexity for the Turing test and knowledge related issues, PAC[20] learning models and so on and so forth. The vast array of examples proposed by Aaronson seems to focus on problems mostly concerning philosophy of science, language and logic. No direct reference to

[18] A NP-hard problem can then not strictly belong to the class NP, but to complexity classes above that, like EXPTIME.

[20] Which stands for Probably Approximatly Correct and indicates a model of human learning proposed by Leslie Valiant (Valiant 2013).

philosophy of mind and cognition is however mentioned, so this leaves a gap that we want to fill. Here we follow the same original approach of Aaronson, and suggest a case of application of computational complexity that is exclusively aimed at cognitive science.

The one example that we will propose is that of tractable cognition (van Rooij 2008; Tsotos 1990; Frixione 2001). This framework is consistent with computationalism and provides a further refinement of it, using computational complexity in order to restrict the set of functions that can be considered as plausible models and explanation for cognitive capacities. Under the premises of computationalism a cognitive capacity will belong to the set of computable functions. As we have already mentioned, no boundaries are here taken into account, and then it is allowed to model a cognitive capacity into a function that has not to meet any running time requirement. This is of course inconsistent with both our common sense and scientific experience. Cognition is, even by an intuitive standpoint, a bounded phenomenon. In order to behave successfully in our environment, we have to produce, actions that have to comply with time and space requirement.

This is also consistent with evidence coming from the experimental findings about the implementation and evolution of cognitive systems. These findings show that there are no evidences of the human brain being a special piece of machinery. It doesn't manifests any unique anatomical features in respect of our mammals closest mammal cousins nervous systems (Roth 2012). Furthermore, the human brain seems to be neither the biggest in absolute terms nor in relation to body mass (Jerison 1973; Jerison 1991). What, however, surely has is "a relatively thick cortex and a medium neuronal packing density found in hominids, humans have the highest number of cortical neurons found in mammals (and animals), which, however, is only slightly above the numbers found in cetaceans and elephants" (Roth, 2012, p. 180). So the specific feature that seems to emerge is the much higher cortical

information processing capabilities that comes from high axonal conductional velocity [21] and a short inter-neuronal distance (because of the human particularly high neuronal packing density (Ibidem)). A feature such as this seems however to indicate at best a performance advantage for the human cognitive system, not supporting the hypothesis humans and their brains have any kind of intrinsically special feature. So, if the human brain is a biological machine that's characterized by certain performance characteristics, it seems reasonable to ask ourselves if these performances are indeed somehow limited, or if, instead we should continue to the brain as some kind of "impossibility engine" (Cherniak 1990). The first of these options is actually called the "bounded brain hypothesis" (Cherniak 1990; Marois and Ivanoff 2005) and it supports the idea that the human brain only has a limited reservoir of processing resources at its disposal. For these reasons the human brain seems to be constrained by the very characteristics of the its material implementation[22] (Simon 1990; Cherniak 1990), but how can we translate such boundaries in a the computational framework?

This calls for a refinement in the computationalist framework and such improvement comes from introducing a restriction on the functions that are considered as cognitively plausible. The threshold that works the distinction between “right” functions and “wrong” function for modelling cognitive capacities is here computational complexity. This requirement is not however exclusive on effectivity[23] but inclusive. It implements a restriction on the previous one provided by effectivity. The application of this threshold to cognitive capacities happens through the following general thesis:

[21] That, on the other hand, comes from the thick myelin sheath that distinguishes the human nervous system (Mark A Changizi 2001; M A Changizi 2007).

[22] In particular, the relation between the total cortical sheet area and the mean cortical synapse density shows that neurons have at their disposal only a limited space of grey area eligible for connectivity.

[23] The fact that a function actually models the target cognitive capacity at all.

(General) Tractable cognition thesis: the set of the cognitive functions CF is included in that of tractable functions.

This thesis, as stated above, is however actually too general to be useful. Without deciding on where to draw the line between tractability and intractability, tractable cognition poses more problem than those that it solves. One first solution proposed is to set the threshold for cognitive functions according to the classic polynomial time rule (Frixione 2001; van Rooij 2008). This way the following thesis came to be:

P-cognition: the set of the cognitive functions CF is included in that of functions that runs in polynomial time P.

This improvement on the general version of the TCT fix cognitive feasibility with polynomial time. This mainly has the consequence of making every cognitive capacity that cannot be modelled into function that belongs in P problematic. Unfortunately, this is not rare. Computational problems that are believed to be at the centre of crucial human cognitive capacities like abductive inference (Bylander et al. 1991), bayesian inference (Chater, Tenenbaum, and Yuille 2006), visual search (Tsotos 1990) and so on[27] have been shown to be NP-hard. This leaves the supporters of tractable cognition with two options: stick with P-cognition and rely on other way of reducing the running time through approximation (Valiant 2013; Chater et al. 2003; Thagard and Verbeurgt 1998) or heuristics (G Gigerenzer 2008; Martignon and Hoffrage 2002; Gerd Gigerenzer, Hertwig, and Pachur 2011), or relax the tractability threshold in order to accommodate more candidate functions. Since we are here interested in the complexity notion itself, rather than in the

[27] For a full list of computational-level theories of cognitive capacities that are believed to be NP-hard see van Rooij (2008, pp. 955-956)

approaches that try to comply with it, we will explore the second option and look at the proposal advanced by Iris van Rooij (van Rooij 2008; van Rooij and Wareham 2007). The solution the she proposes is to let the polynomial time requirement go and apply in its place another tractability threshold that comes from a refinement of computational complexity called parameterized complexity (R G Downey and Fellows 1999). This theory stems from the observation that certain NP-hard functions have a polynomial running time when a subset of the input is considered. This means that, if a certain number and type of restrictions is applied to the whole input size certain, otherwise intractable functions, can be considered tractable when the right conditions are met. The small portion of the input that has a non-polynomial running time is called a parameter, a function that run in polynomial time for the restricted input size takes the name of fixed-parameter tractable and the relative complexity class will be FPT. Without the need to go into further details, we have all the ingredients to state the following thesis:

FPT-cognition: the set of the cognitive functions CF is included in that of functions that runs in FPT-time.

This relaxation on the original P-cognition makes possible to account to a wider range of possible computable functions in order to explain cognitive capacities and phenomena that, otherwise, would call a theory revision or, worse, a complete theory rejection. Even if some elements of the various forms of tractable cognition are hinted it still not clearly stated in the literature what kind of theoretical role does tractable cognition play inside the interdisciplinary endeavour of cognitive science. We will provide with a take on this issue in the following paragraph.

2.3 Tractable cognition as a plausibility notion

At the end of the previous chapter we indicated a number of feature that are commonly associate with complexity. Here we will not make the case for computational complexity possessing them all of course. We will instead support the idea that even the possession of some of them still indicate the possibility of fruitfully applying the notion for genuine philosophical purposes. Some of these aspects are already captured by the complexity notion itself, and some other are instead better grasped by its application into cognitive science. It's on the capacity of a certain notion of complexity of improving on theoretical clarity and depth that philosophical relevance stands. Here we will evaluate the advantages and disadvantages of both computational complexity and tractable cognition.

The first feature of complexity to take into consideration is its association with "threshold" arguments that are particularly frequent in criticisms against the reduction of consciousness: the "complex enough" case. Computational complexity seems to be pretty well equipped from the start to deal with such ambiguous cases. Differently from other complexity measures, like Shannon information for example, computational complexity provides an indirect measure and not a direct one. Complexity classes are used to "organize" functions into broad running time (for time complexity) categories and then no actual direct measurement takes place. This should make computational complexity particularly suited to capture semantic notions of complexity, rather than the syntactic ones where capacity of measuring length and size seems to be crucial. Also, computational complexity comes already equipped with its own threshold. This can be applied through equivalence to the desired phenomenon. Tractable cognition is a clear example of this versatility.

Since computational complexity is indeed a formal notion, it also improves on the clarity of the notion of philosophical complexity at which it can be applied. This of course happens when computational complexity is taken outside its original application and stretched in order to provide a formal index for characterizing otherwise ambiguous theoretical notions. The case of tractable cognition is here paramount. It allows to bridge computational complexity with the idea of bounded cognition and provides then a way of characterizing the distinction between plausible explanations and models of cognition and implausible one. This notion of "plausibility" that we just used has been already mentioned d before in the text and should be considered as a requirement for explanations and models. The requirement is relative to the actual capacity of such explanations and models to express only the necessary and sufficient proprieties that are needed in order to perform an actual explanatory role. This notion should be balanced to that of "richness", being the actual amount of details that an explanation of a model actually implement. According to Gervais and Weber (2012) these two aspects should be well though and balanced in order to get an effective explanation[30]. Given the feature of computational complexity and tractable cognition that we have discussed above, it is possible to see how plausibility could be linked with them in order to get the following thesis:

Plausible cognition thesis: if a candidate explanation for a cognitive capacity can be modelled into a tractable (P or FPT) function, then such explanation should be considered also cognitively plausible.

Linking plausibility with computational complexity and especially tractable cognition provides philosophy with an addition tool to evaluate theories and

[30] The authors main target is here mechanicism, but the general considerations about plausibility still stand even in relation to our topic.

the models that are there implied. The possibility to variate the tractability threshold accounts also for the need of balancing between cognitive plausibility and richness, that here takes the form of the common-sense evidence about the capabilities of human cognition. All these possibilities definitely account for the philosophical relevance of computational complexity and even more for that of tractable cognition. FPT-cognition in particular adds a further dimension of philosophical enquiry, due to the fact that the intuition for problem parameterization is often taken from theories and commons sense. This way a further (philosophical) plausibility evaluation takes place, corroborating the hypothesis that computational complexity can be of philosophical interest.

Chapter 3

Understanding simplicity through complexity

The goal of the previous chapter was to show how stratified and intricate of a notion complexity can be. In the end of it we made a choice among the various measures of complexity that have been proposed and explained why computational complexity and the cognitive theories that have been built onto it have indeed a philosophical value. We also proposed that CC, when utilized trough the TCC, takes the form of a plausibility notion that, when implemented, can increase the likelihood and accuracy of theories about cognition.

In this chapter, we will begin testing this hypothesis by considering in details one of the most important relationship that complexity entertains: that with the notion of simplicity. While it may seem paradoxical to talk about simplicity in order to clarify complexity, the link between these two notions is indeed a strong. The fact that "complex" and "simple" are somewhat related is intuitively evident but this relationship has been also strongly investigated throughout the philosophical and scientific history (Weaver 1948; A. Baker 2013). Minimizing complexity while at the same time maximizing simplicity has often been supported as one of the main good practice in science. That is why in philosophy of science the notion of simplicity takes the form of a principle for evaluating theories. This alone should explain our interest on it, since it seems to perform the same role that is often recognized to complexity and that we are attributing to computational complexity in cognitive science. Besides that, it happens to be that simplicity in philosophy of science is articulated, as we will see, in two aspects that are respectively tied to various

measures of complexity. This makes of simplicity also an internally "complex" notion in itself. Investigating this particular relationship between the two notions of will also open us the opportunity to take the analysis into applications in cognitive science. That will be done by critically considering the hypothesis that claims that the behaviour of cognitive agents is inherently driven, or determined, by the implicit use of a principle of simplicity. Such a proposal is indeed quite ubiquitous in cognitive science if we accept an intuitive link between simplicity and the notions of parsimony and economy. However, a more precise and technically savvy proposal have been advanced by Chater and Vitànyi (2003; Vitanyi 1998). There the authors defend the idea that the principle of simplicity that is behind the determination of the behaviour of cognitive agents is ultimately based upon the notion of Kolmogorov complexity. This has two consequences: for starters, it grounds simplicity on a notion of complexity that, as we have seen, has descriptive and syntactic purposes; furthermore, it promotes an absolute way of intending the simplicity principle according to which cognitive agents should always prefer the simplest solutions.

By contrast we argue that such a view on simplicity, while compelling for some aspects, is not complete nor accurate enough to capture the theoretical depth suggested by the philosophical history of the notion. We propose then that for notion of simplicity to be used coherently and fruitfully in cognitive science this notion needs to be implemented so that also also a form of cognitive parsimony can be taken into account. To reach such a redefinition we have to establish how a principle of simplicity that started a criterion for the selection for theories can be translated in a cognitively effective principle about the selections of problems.

In order to understand what are the philosophical implications at work will initially consider how the development the notion of simplicity had in philosophy of science. We will see then how simplicity has been considered in

philosophy of science as a virtue of every good theory (scientific, philosophic, mathematic and so on and so forth) and how this notion articulates in the form of a principle that articulates in two basic components: elegance and parsimony. To individuate the points in which simplicity is in need for a revision we will compare the conclusions taken from the philosophical analysis with the particular explanatory requirements that the study of cognition requires. This way we will see how parsimony and elegance are indeed still relevant components of a principle that is tailored on cognitive systems instead of science as a whole.

After seeing in details how the accounts from Vitaniy and Chaitin unfolds, we will propose that computational complexity can provide the necessary theoretical tools for updating the principle of simplicity that they support. By doing so, we argue, it is possible to reach a more complete picture of the different aspects of the simplicity principle and, also, to introduce into a plausibility threshold into it so that it can be rephrased as follow: cognitive agents do not prefer the simplest solution among all the possible ones, but they prefer the simplest solution among the most plausible one. Where for plausible we intend those solutions that can be achieved in practice.

3.1 Simplicity in philosophy of science

In philosophy and science, the simplicity notion indeed has a long history and, traditionally, it simplicity take the form of a virtue that good theories should possess in order to be considered explanatorily sound[31] (Gauch 2003, pp. 270-277). According to this then we can then say that not only simple theories, but also simple explanations, simple solutions and simple demonstrations should be preferred in place of more complex

[31] For example, we can find traces of it in Aristotle's *lex parsimonie* (Posterior analytics), in St. Thomas Aquinas (Hoffmann, Minkin, and Carpenter 1997), Galileo Galilei, Immanuel Kant and Isaac Newton *(A. Baker 2013)*.

alternatives. Simplicity as intended becomes an index that synthetically express the possession of those properties that are listed as reasonable hints of the presumed plausibility of a certain theory, or even of its truth. (Swinburne 2001). For the purposes of our analysis we have, however, to reach a way more precise characterization of the notion at hand. This means overcoming the more intuitive aspects and make a step toward more precise accounts. It will be then paramount to investigate what are the specific factors that make a theory simpler (and by consequence less complex) than another one and how such factors can be investigated in a systematic way.

More information in this sense come if we continue in our historic overview. One of the most obvious, and famous, explicit reference that we can find of simplicity is represented by Occam's Razor (OR) (Wright 1991, pp. 77-104). What is important to note is that OR takes the form of a virtuous principle that applies as an ethical criterion for making good and valuable theoretical work. Abiding to this rule lead, in philosophy, for the inherent preference for those theories that explain a certain phenomenon with the least possible ontological commitment. When conceived like this OR takes the shape of an a priori ontological principle. For better explaining this point we should evaluate how OR is applied in the actual philosophical debate. Here we will take the two influent theories of cartesian dualism and materialist monism as a case study. Both these theories have the mind-body problem has their explanatory target, however they widely differ in the way in which they explain the phenomenon. While cartesian dualism commits to two different metaphysical substances (the *res cogitans* and *res extensa*,), materialism monism only commits and make use of a single material substance. Following OR we can then observe how dualism is indeed more ontologically committed and the, more complex than less burdensome than its monistic alternative. This comes from the fact that, in order to provide an explanation, dualism actually postulates a world that is more complex than the one

represented by monism. A world, with all the laws that it expresses, in which we have to find evidence for two substance instead than one. According to this evaluation, that for the time being we can call of ontological parsimony, monism should be *ceteris paribus* preferred a priori because it inherently expresses le complexity and, then, it is more simple.

Through this example we can clearly see how OR can be utilized and in what sense it takes the shape of a ontological criterion of parsimony. This makes less vague the nature of the "evaluation" that takes place when the principle of simplicity is applied. However, it also allows us to ask ourselves id this kind of ontological parsimony is exhaustive, or if there are instead some further way in which simplicity can be philosophically characterized. In order to clear this point, it will be useful to again rely on philosophy of science and look at how the philosophical debate handled this issue. In this field two different aspects of the simplicity principle have been in fact opportunely distinguished and this makes the question about the simplicity of theories twofold: on the one side simplicity is inversely proportional to the complexity that derives from the number of hypothesis that a theory needs in order to provide an explanation; on the other hand simplicity is also inversely proportional to the complexity that derives from the number of ontological entities that a theory postulates (this reflects the one that we have encountered before). Such elements constitute key aspects of simplicity and have been named respectively elegance (or also syntactic simplicity) and parsimony (or also ontological simplicity). According to this distinction a theory can be elegant, or syntactically simple, is it provides an explanation that is concise, without making use of any superfluous element in its expression beside the strictly necessary ones. Instead, the parsimony, or ontological simplicity, of a theory is function of the types and number of entities that it postulates. This distinction has the consequence of improving the granularity of the simplicity evaluation. A theory can be elegant while at

the same time being not parsimonious and vice versa. Another effect of this distinction is that of making the principle of simplicity a global index that can be assessed on the base on local properties of the theory at hand. A characteristic this one that makes it pretty similar to the informal notions of complexity that we have analysed in chapter one. Acknowledging this refinement not only brings us some reference to the way in which have already treated the notion of complexity, but provides us with also a refinement that updates our understanding of simplicity. We can now say that a theory can be considered simple if it achieves the best compromise[32] between a concise expression and a parsimonious representation of the world.

Going further we can give a look at how this distinction can suggest us further ways in the phenomena of simplicity can be characterized and analysed. First and foremost, the relationship between simplicity and complexity reveals that the first and the last are intimately linked also in relation to elegance and parsimony. If the elegance of a theory is determined by its syntactical simplicity, and such simplicity is inversely proportional to a form of complexity, it is also possible to say that a measure of elegance will also be measure of complexity in its own right. A measure that is grounded in a form of syntactic complexity that has been defined in several ways in the literature. Elliott Sober (2002, p. 2), for example, proposes that such complexity should be based on the number of symbols in which the statement of a theory is expressed. In the same direction goes the Minimum Description Length principle of Jorma Rissanen (1978; Griinwald 2005). Such principle constitutes a formalization of elegance that utilizes the idea of knowledge as data compression (and then Kolmogorov complexity) as a cornerstone for the

[32] Here we phrase simplicity as a compromise because it may also be the case that to the most concise and elegant formulation doesn't correspond the most parsimonious one. The two notions of elegance and parsimony may also be in contrast. It can then happen that the postulation or more entities may lead to a good explanation, like in the case of the discovery of the planet Neptune by Le Verrier and Galle.

measure of simplicity itself. One last measure of elegance is Nelson Goodman's logical testability (Goodman 1943; Goodman 1955; Goodman 1958; Goodman 1959), for example, measures the number of logically superfluous statement in a theory, where superfluity is determined through logical properties.

For the above-mentioned ways of formally defining elegance the following observation holds: such a characterization of elegance, both in its informal and formal versions, closely binds this form of simplicity to the language in which the theory is actually stated. This has the consequence of making impossible, or at least difficult, the comparison of theories that are expresses in different languages, regardless of them being formal or informal. The fact that elegance cannot be relieved from this deep dependence with the language of the theory[33] highlights even better how important of a role parsimony plays. Likely to what happened for elegance, also parsimony can, and should, go through some further technical refinement. Some hint from where to start come from our previous thoughts on OR and, especially, ontological parsimony. According to that preliminary depiction of parsimony parsimonious theories are those that take into the equation less entities in comparison to their more complex equivalents. Following this way of thinking, parsimony refers to what the theory says more than how it is expresses. One way to render parsimony into a measurement is to think about it probabilistically. If two theories can be reduced into two statements and these end up being equiprobable, then it can also be said that the original theories have indeed the same semantic complexity and posit the same type of simplicity (Sober 2002). However, Elliott Sober's is not the only one available. Charles Peirce's testability principle (Peirce 1931), Karl Popper's falsifiability (Popper 1959)and John Kemeny's testability (1955) can all be

[33] In fact the definition of elegance in terms of syntactical complexity makes such a relief impossible.

thought as way of characterizing how easy is a certain theory to test and corroborate. Parsimony may indeed follow from these types of measure because ontologically simpler theories will also be easier to test and corroborate.

What should be clear at this stage is that if parsimony has a particular characteristic is that of being much more resistance to attempts of formal reduction. That comes from the target itself of this measure. Formally accounting for parsimony, like said before, means to render in a measure form the what the theory means and what kind of consequences this has on the quality of the explanation that it gives. However, even if a completely satisfying measure of parsimony cannot be individuated right now, we sure have a better picture about how elegance and parsimony mutually interact to form the synthetic formulation og simplicity. A theory can be considered simple, or simpler that another one, when it is concisely formulated (elegance) and it also postulate only the necessary and sufficient number of entities (parsimony).

Another thing that should be considered about simplicity is that, while the use of this principle can be well justified in philosophy of science, that may not be the case for other disciplines. This kind of considerations brings about the topic of the globality of the simplicity principle (Crick 1990; Sober 1990) as we have stated it. The application of the same principle outside of its natural environment may indeed need some tweaking. Especially because not every field has theories, explanations, entities and formulations as their basic components. However, while these components may actually vary, the taxonomy of simplicity, its relationship with complexity and the measurement factor can be translated to fit different theoretical targets. After this detailed explanation of simplicity, it is time to move to our original field and look at how this notion, with all its ancillaries, can be extended to cognitive science.

3.2 Simplicity and cognitive elegance

Various proposals to suggested that simplicity, rather than being only a methodological principle, may indeed be a fundamental component principle that guides the behaviour of living organisms (Mach 1914). What these proposals suggest is that organisms, thanks to intrinsic characteristics of their cognitive systems, apply a simplicity principle when they execute a task (Chater 1997). In light of the analysis that we have done in the previous paragraph, it is reasonable to apply the same dual interpretation of simplicity also for the application of this notion in cognitive science. This way of thinking suggests us that this kind of "cognitive simplicity" could be subdivided into the two aspects of elegance and parsimony. However, before starting this kind of analysis, we have to set the ground and specify what are the scopes and peculiarities of cognitive science.

First of all, cognitive science is concerned with explaining cognition, the elements that compose it[34], namely cognitive systems. A cognitive systems can be defined as "a dynamic order of parts and processes standing in mutual interaction (Bertalanffy 1968)" that possess the necessary and sufficient characteristics to express phenomena that can be ascribed to cognition. One of the most widely accepted accounts about the nature of cognitive systems is the information processing one (Neisser 1967). A cognitive system should then be interpreted as a processor that perform a cognitive task by receiving information (input) from the external world, process it according to several internal rules, and outputs a behavioural content. This functional definition can also be restated so that it can fit not only global interpretations like the one above, but also the single capacities and sub-capacity that make possible

[34] Here we assume naturalism and reductionism as a premise (Nannini 2007; Daniel C. Dennett 1991; Daniel C. Dennett 1995). However, this approach is of course still debated in philosophy (J. Fodor 1975; McGinn 1999; Dreyfus 1972), so we are not suggesting that this is the only option available.

for a system the performance of a cognitive task. This makes of course easy to make the transition from capacities to functions and say that every cognitive capacity is defined by the input that it gets and the output that it has to provides (Cummins 2000). This also qualifies cognitive capacities and sub-capacities as functions, encouraging the computational explanation that permits to consider them as computational problems (van Rooij 2008). All of this has however to be summed up with some intrinsic characteristics of cognitive systems: they are epistemically limited by definition[35], they are not concerned with general theories but with problems and solutions that are inherent in the capacities that they perform.

These facts about cognition and cognitive science adds to the fact that is now possible to reinterpret simplicity in a cognitive fashion without fear of being too dispersive. If a living organism can be considered a cognitive system, and is then able to express cognition through its cognitive capacities, it is also possible argue that for simplicity to have a part in this scheme then it must have an influence on the way in which systems perform. These considerations provide hints on how simplicity needs to be reconsidered in order to fit into cognition. Cognitive systems intrinsic epistemic limitations we have to conclude that the simplicity notion there at work cannot have the be scope of the one that has been tailored for discriminating between general theories. A prospective "cognitive simplicity" should instead be able to handle problems and solutions that are tied with the relative cognitive capacity and are then limited in scope. Furthermore, the same notion should be tailored down so that its goal not the best possible theory, a but instead the solution that's good enough. This because the solutions that are implement in and by cognitive systems are not always similar to theories. In the majority of case they look more like shortcuts and heuristics to which it is difficult to

[35] They possess a limited point of view on the external world and then they access only a part of the information present there.

apply an absolute criterion of preference of (Daniel C. Dennett 1995; Daniel C. Dennett 1991). If, then, a notion of simplicity can be translated into an effective cognitive principle, it should be able to account for the possibility of choosing (not necessarily consciously) only among the solutions that are really accessible to the cognitive system at hand and not among all the possible solutions to a given problem[36].

One attempt of satisfying even a part of these requisites has been proposed by Nick Chater and Paul Vitanyi (1999; and Vitányi 2003). According to their thesis a cognitive system naturally prefers simple explanation and solution whenever it performs a general cognitive task because of the actual way in which information is ordered and codified. It through this preference simplicity that cognitive systems are able to solve the problem of induction[37], and then select only the relevant part of the available information. Chater and Vitanyi's solution stands in considering simplicity as a fundamental property of the way in which cognitive systems interpret the external world into patterns. This is also why this proposal also grounds simplicity on a measure and, especially a measure of complexity. Without a measure, it would indeed have been difficult to provide an explanation on how the preference is awarded. The measure that the authors use as reference is that of the compression of data. A notion of complexity that, as we have seen in chapter one, is closely tied to that of Kolmogorov complexity. This measure provides an indication on how the data in a set can be compressed (Grunwald and Vitanyi 2003; Vitányi and Li 2000), and then on the possibility of them to be arranged in the most concise way and then easier to computationally retrieve. According to this way of intending simplicity a cognitive system will always prefer those information patters that allow the

[36] That correctly models a cognitive capacity.

[37] This problem is based on the fact that every model is always compatible to every possible finite set of data. This makes models always underspecified.

most compact encoding in respect to the data (Chater and Vitányi 2003, p. 20). If we take now a searching task in a perceptive field as an example, a cognitive system will process only the relevant information selecting through the perceptive spectrum and preferring those patterns that can be better encoded and compressed.

However, beyond the description of the principle itself there is also another remark that needs to be done. Chater and Vitanyi's account on simplicity is indeed a way of individuating how simplicity can be involved in the selection of solutions by cognitive systems. However, in the first part of this chapter we have also described how simplicity can be divided into two different, and not always symmetrical, aspects: elegance and parsimony. By applying this distinction here, we can see that the kind of cognitive simplicity, while cognitively sound, is not complete or exhaustive. On the one hand, it utilizes the one of the measures of complexity that has been the most associated with formal definitions of elegance. On the other hand, the formulation itself favours the way in which information is encoded and, then, its expression or extension. It should be concluded than that this kind of cognitive simplicity resembles more to elegance, rather than a complete and notion of complexity. What about parsimony then? Is it possible to individuate a version of this notion that is cognitively sound? In the next paragraph, we will provide an answer to this question and present our proposal for a cognitive parsimony.

3.3 Towards a redefinition of parsimony

Like we saw in the first chapter the parsimony is function of the number of entities that a theory needs in order to provide and explanation. It addresses then the semantic complexity of a theory, leaving compactness to elegance or syntactic complexity. Thanks to Chater and Vitanyi's proposal it

has been possible to at least individuate one possible example of how simplicity can be applied in cognitive science. While the requisites imposed by the study of cognition have been partially defined, it is possible to provide an integration so that a more exhaustive account can be reached. Cognitive systems epistemic limitation is not only expressed by the fact that they have only a partial access to the external world, but also by the fact that they have to comply both to ecological and constitutive constraints. The first are determined by the context in which the system is immersed and the task that it should perform. Shooting an arrow to a target, drinking a glass of water, answering a phone call or opening a door are all tasks that need to be performed in a timely manner to be successful. Conversely, the cognitive system has to comply with the fact that both its architecture and its implementation also impose serious constraints. Shooting an arrow can be effectively undoable if someone does not possess the strength for drawing the bow, or it is too tired to do that. Their concrete realization of cognitive systems makes then cognitive systems intrinsically resource bounded and this is in itself another side of their limitation. Acknowledging this aspect provides us with means rectify the simplicity principle: a cognitive system does not only have a preference for elegant information, but that it also applies a parsimony evaluation. This cognitive version of parsimony would not be a property of information, but it would instead be based on the resources that a cognitive system has available and to the requisites of the target problem. It will measure not the intrinsic properties of how information appears or can be encoded, but instead will try to account for the difficulty of a cognitive task. Of course, to do so it is crucial to clarify the nature of such resources and, therefore, approach a rigorous and clear way or measuring them while at the same time providing a threshold for determining where the plausibility limit should stand. This way we should be

able to reach a simplicity notion that is cognitively plausible and while at the same time being also philosophically complete.

For capturing cognitive parsimony has we have just defined we will again entrust computational complexity theory. In chapter one we have seen that this theory is a measure of the computational resources that a problem requires to be solved in function of the size of the input. As previously remarked, the application of this theory to parsimony requires of course to accept the possibility of computationally explaining cognition. However, doing so permits us to have access to a possible bridge between the resources of a cognitive systems and the computational resources of time and space. If cognitive systems are computational than it is possible to equate the problem that they solve to computational problems. This is especially true if the TCT, with its equation between computational tractability and cognitive plausibility, is taken into the equation. Doing so computational complexity definitely provides a threshold for distinguishing between those problems that are computationally tractable, and those that instead are computationally intractable. How such application works becomes clear when we consider an example. If we take again our search task form before, a cognitive system will not only apply a preference for patterns of data, but it will also have to evaluate the actual processing weight of those data and how this influences the intrinsic difficulty of the search task. Since not all the computationally tractable cognitive functions will be accessible from the start to a cognitive system, the choice among the various solutions will happen among those that are available. Also, because computational complexity is not a direct measure, but asymptotic, the preference for parsimony will not be absolute, but instead only relative to tractability. This way a plausibility threshold will be applied to contingent choice that does not apply to general theories but to local solutions.

According to these results we can now try to formulate a definition of parsimony that makes good use of tractable cognition. Again, like happened with OR and cognitive elegance, we will need to translate a notion into a principle of preference. By doing so we obtain this: in a finite set of accessible solutions a cognitive system will prefer the solution that is takes less time and less space and, therefore, is less cognitively difficult. This version of parsimony is indeed similar to the philosophy of science one. The number and the complexity of the entities being the defining factor that guide the choice between alternatives.

For concluding we can now try a synthesis between cognitive elegance and the cognitive parsimony that we just proposed. A cognitively plausible and philosophically complete notion of simplicity is defined as that notion that guides the following methodological principle: a cognitive system will prefer, select and promote simple processes/solutions because they permit to express information in a concise way (cognitive elegance), and also because through such processes/solutions the task at hand can be practically performed in a timely manner (cognitive parsimony). While cognitive elegance will account for phenomena like saliency and information selection in the environment, cognitive parsimony introduces pragmatic limits that accounts in the intrinsic limits of cognitive systems. Like for their eminently philosophical counterparts the efficacy of both this aspect of cognitive simplicity stands in the possibility of reaching the right balance and compromise. A compromise that is indeed possible especially if we look at how the two measures of complexity adopter, Kolmogorov complexity and computational complexity, can partially overlap on some complexity classes (Fortnow 2004). The fact that an overlapping exists, instead of a complete identity is another point that corroborates how simplicity also in its cognitive form can be separated into to concurrent, but not coincident, aspects. While non-exhaustive, these considerations provide a clarification of how

computational complexity fits inside different roles and adapts to goals that are different from the one originally thought. We started from a structured philosophical notion for evaluating theories and we ended with one that can be applied to single systems. The application needed some philosophical fiddling, but that is exactly the drive behind this thesis and confirms how theoretical thinking can be useful to fit technical notion into the wider scope of main frameworks. In the next chapter, we will go back and talk again about theories and we will have the opportunity to check how frameworks can be evaluated from the notion of complexity that they use.

Chapter 4

Confronting dynamical and computational complexities.

In the first chapter, we introduced the Dynamic Hypothesis (DH) as one of the principal theoretical consumer of complexity. We have also have seen how DH has often been presented and characterized as one of the principal alternatives to the widely popular Computational Hypothesis (CH) in cognitive science. This is also reflected, at least intuitively, in the notion of complexity that this accounts expresses. What we are going to do in this chapter is considering in detail the type of complexity that DH shows. Then we will propose that while the theoretical distance that separates these two approaches may seem to be significant, there are good reasons, we argue, for reconsidering the nature of the relationship between the dynamical and computational ways of understanding cognition. This goal will be reached mainly through the claim that CH and DH, rather than being competitors, are complementary framework in the explanation of cognition. This, we suggest, becomes evident when the two different notions of complexity that these theories use are considered. What emerges from such analysis is that while these two notions of complexity can be different, they are not contradictory. DH, we recognize, uses a notion of complexity derived from dynamical systems theory that seems to points toward psychological plausibility, or richness. On the other hand, CH may appeal to a computational notion of complexity that introduces elements of cognitive plausibility in the theoretical framework of cognitive science, or plausibility tout court. In the following paragraphs, in order to test this last hypothesis, we will proceed as follows:

First we will consider how the debate between dynamicists and computationalists, started in the first place and what are the main features of DH and of its notion of complexity; Then we willl see how CH evolved and consider how complexity can be recognized inside the CH framework; following that we will take a look at how computational complexity applies to cognition by linking computational tractability with cognitive plausibility; In the end we will argue that if we take the two different notions of complexity by comparison there are elements to consider them not adversary on the same ground, but as complementary model of explanation.

4.1 Dynamicism in details

The computational theory of mind, or Computational Hypothesis (CH from now on), has often been regarded as the received view in cognitive science. That comes from the fact that it provided, through its analogy between the working principles of the mind/brain and that of artificial computers, one of the very first tangible model of cognition. Instead of considering the riddle of criticisms[38] that it has received during history of cognitive science, we will concentrate on the debate between the supporters of CH and those philosophers and scientists that have proposed that a cognitive system should be intended as a Dynamical System (DH).

The promoters of DH start from the premises that understanding natural cognitive system as dynamical models is radically different from considering them as computational systems. However it should be clarified that DH supporters are not interested in criticizing any version of CH, but they often concentrate on that hypothesis that has been called the "paradigm of the computer" (PoC) (Cordeschi and Frixione 2007). The PoC could be defined as

[38] In this paper we will not address the following critical points: the semantics/syntax distinction (Putnam 1960; Searle 1980; Searle 1992) or the computability tout-court of cognitive processes (Penrose 1999).

a pretty restrictive take on computationalism that can be synthesized by following statement: a natural cognitive system works following the same principles of digital computers. This is clearly restrictive, since it presupposes that computationalism always intends cognitive systems as Von Neumanesque general-purpose architectures. However, what we actually want to do is evaluate DH by considering the features of the explanatory model that it proposes and not in relation to how it defines its polemical target. So, for now we will concentrate on those elements that introduced genuine novelty in the cognitivistic debate.

First it should be noted that DH is actually a twofold hypothesis (T van Gelder 1998) about the nature of cognitive system. The first weaker sense in which DH can be interpreted is called the "nature hypothesis" and claims that cognitive agents actually instantiate dynamical systems in those parts that are considered to be responsible for cognitive performances. The second takes the name of "knowledge hypothesis" and claims that natural cognitive systems[39] should not only be considered as dynamical systems, but they should also be modelled after dynamical systems. These two interpretations of the term dynamicism both make use of the notion of dynamical system. The notion of dynamical system, as often happens in cognitive science, is a borrowed one and it comes from mathematics and science where it has already found numerous uses and definitions[40]. However, for what matters cognitive science, is often considered enough to say that dynamical systems are state-determined systems "with numerical states that evolve over time according to some rule" (Van Gelder and Port 1998, p. 5). Where for "state-determined" is intended that the current state (the set of variables) happens inside a state

[39] Since artificial cognitive system may well be computational, DH is a theoretical hypothesis only over the nature of cognitive systems that are found in nature and that have therefore naturally evolved.

[40] See van Gelder (1998, p.618) for a series of examples of dynamical systems in physics and mathematics.

pace[41] (or phase space) and that this state determines, given an evolution rule, the future states in which that system may evolve in time. This means that natural cognitive system and the phenomena that are ascribed to them should be intended and modelled utilizing the same concepts used for describing the growth of a population of bacteria, the behaviour of an undamped pendulum or that of the solar system. There is no need for a sharp distinction between different processes, what's really important is the rate of change of the states of the system and the geometrical relations that develop in the state space. Everything cognitive has to be conceived, according to DH, as globally conceived and inherently relational. Systems are intended as a set of mutually bonded variables that evolve together in space (the environment) and time (they evolve simultaneously). These background features of the notion o dynamical systems have deep philosophical consequences and give to DH a set of very unique characteristics:

Emphasis on temporal evolution: dynamical systems' processes are not modelled around a notion of discrete state succession, but it emphasizes the rate of changes of the variable of the system in the unit of time. It's not the order to be important, but the changing aspects of the mutually linked variables. At the same time this approach enables to consider cognitive systems not as a diachronic architecture, but as system that instantiate processes always working simultaneously and in real time;

Emphasis on holism: with no distinction between central and peripheral processors, or between the brain and the mind, DH is a naturally embedded and embodied model of cognition (Thelen 1995). This means that, while there is no longer the need to sharply distinguish the mind from the brain and body, there's also no need to strongly separate cognitive systems from their environment. For these reason we can also consider DH as a naturally

[41] The space where we find all the possible state in which the system may be in.

extended theory about cognition (Clark and Chalmers 1998). Furthermore, a distributed, parallel and non-modular model of the mind can be easily conceived under DH, since there is no use for the notion of symbol and, therefore, for that of representation. These same features are responsible of strongly binding DH with a notion of complexity. In the following paragraph, we will go through the reason why it is so and what are the main characteristics of such a notion.

4.2 Complexity in the dynamicist hypothesis

We've already seen how DH re-imagines cognition as a continuous, interactive and geometrically organized phenomenon. What is achieved this way is a theoretical hypothesis that naturally conceives cognition as a complex phenomenon and cognitive systems as inherently complex systems. DH does so thanks to the relation between dynamic systems and dynamic system theory[42], but also because under DH cognitive systems are considered as topological structures in which "the interaction among constituents of the system, and the interaction between the system and its environment, are of such a nature that the system as a whole cannot be fully understood simply by analysing its components" (Byrne 1998, p. viii).

DH surely inherently presupposes a notion of complexity thanks to its emphasis on time evolution. In the real world, decisions have to be taken in split seconds and actions, as grabbing a cup and taking a sip of water, are needed to start and stop at just the right moment in order to succeed. All of that also seamlessly happens in complex continuum where a number of multiple processes all take place simultaneously and on different time-scale[43].

[42] A branch of pure mathematics concerned with the behaviour of complex systems. (Alligood, Sauer, and Yorke 1997)

[43] Since there is not a discrete state subdivision different processes may well occur simultaneously but with different rates of change in the phase space.

However, DH doesn't only pose an accent over the necessity of intending cognitive phenomena as real-time processes, but it introduces also the necessity of interpreting cognitive systems as real living beings acting in a living real world. The same actions and decisions of before are not taken by idealistic systems or architectures. Furthermore, rather than showing a computer-like architecture, they are organized in a net-like structure of interdependent elements which apparently doesn't behaves like any Von Neumann architecture we know about. As a matter of fact, neuromorphic models of information processing, such as the artificial neural networks of connectionists, are considered by dynamicists example of dynamical systems [44](Tim Van Gelder and Port 1998; T van Gelder 1998; Tim Van Gelder 1995b; Horgan and Tienson 1992). Understanding the CNS as a "single dynamical system with a vast number of state variables" (Tim Van Gelder and Port 1998, p.34) comes pretty natural under DH.

Inside the dynamicist framework there has seen application in the modelling of decision processes (Townsend and Busemeyer 1995) and sensorimotor activity (Saltzman 1995). In linguistics DH has made feasible a semiophysical (Petitot 1992; Thom 1988) approach to some of the trickiest feature of natural language, such as linguistic meaning, compositionality, semio-genesis and lexical polysemy. Especially this last field of enquiry reveals another characteristics of the dynamical framework: qualitative and phenomenological aspects of cognition can be naturalistically explained (Petitot 1999) without the need for a reduction or an elimination. The flexibility of dynamical systems makes at least analogical explanations possible, so "even without an elaborate data time series, one can study a mathematical model which exhibits behaviour that is at least qualitatively

[44] However, while dynamicists are keen to acknowledge artificial neural network as genuine dynamical systems, they often consider connectionism itself as an unfinished attempt at overcoming computationalism (Tim Van Gelder and Port 1998, p.32)

similar to the phenomena being studied" (Tim Van Gelder and Port 1998, p. 16).

In conclusion, we can now safely affirm that complexity is indeed a fundamental element of DH. Cognitive systems qualify as complex systems when considered dynamically because the model is closer to the reality of actual cognitive agents. By doing so DH places cognitive systems in real time and poses it in relation with a naturally complex environment. What complexity does in DH helping at closing the gap between the mind and body. Cognitive phenomena don't have a set of (computational) restraining them and the same goes for the actual architecture of the cognitive systems. Avoiding the constraints of computationalism helps at reconsidering the way in which a cognitive system is said to be cognitive in the first place. A dynamical cognitive system qualifies as cognitive thanks to what it does and not because of what it is. The dynamicist model does not define cognition in top-down way, but instead is the model that has to adapt in order to reach a better understanding of the explanandum. This give to the notion of complexity in dynamicism a quite specific flavour. In DH complexity is introduced not only in order to close the gap between models and reality, but also because there is a clear accent toward maximizing psychological plausibility. The actual constraints and limitations of the cognitive system are somewhat overlooked in order to reach a deeper understanding of those cognitive facts that were previously considered as irreducible or subject to plain elimination. Furthermore, This way of using complexity is a clear plea toward richness (Gervais and Weber 2012), namely the amount of details that a theory has to comply for to be considered an acceptable explanation for a certain phenomenon. On the contrary, computationalism apparently focuses on finding constraints in order to keep the complexity of cognition at check, dynamicism exploits cognition in order to make our explanation of it closer to the actual experience of a mind.

In the following paragraph, we will instead see how complexity works in the computationalist side of cognitive science. In contemporary computationalism complexity surely plays quite an important part but with very precise characteristics.

4.3 The evolution of computationalism

We have seen how criticisms coming from DH supporters focuses on what we have called the PoC version of computationalism. We also said that PoC actually tells a very reductive story about computationalism and in this section, before we get to properly analyse the computational aspects of complexity, we'll first look at how computationalism evolved in response of the numerous arguments and criticisms addressed to it. From what we said earlier about the theoretical background of DH, it's possible to synthetize all the criticisms against CH in the following two criticisms:

1. CH treats cognitive systems as sequential machines. While cognition works in real time there is no indication that computational models of cognition may be able of doing the same;
2. CH treats cognitive systems as they were general-purpose artificial calculator. In fact, a cognitive system is a very specialized architecture dedicated to a number of very specialized task. Furthermore, there seem to be no strong relation between the parallel and net-like structure of the brain and the strictly hierarchical architecture of computational devices.

While apparently different these arguments really point in the same direction. What DH accuses computationalism of is to treat cognition as a much simpler phenomenon that it really is. CH is then considered insufficient for effectively explaining the complexity of genuine cognitive phenomena (Piccinini 2010b).

However, since the definition of the actual notion of computation has deep consequences on the characteristics of the theoretical framework, some of attention should be dedicated at analysing this notion and its evolution.

The first accounts of computationalism (Putnam 1967) actually traced a strong analogy between the state of a cognitive system and the state of a Turing Machine that also survived in computational-representational theory of cognition (J. Fodor 1975). These theories surely suffered from claiming a bit too much about the nature of cognition. While Turing Machine is currently the best model of computation that we can account for on the general level, it may not be the best model for describing the kind of computations that can be found in the brain. Connectionist models (McLaughlin 2003; Rumelhart and McClelland 1986; McCulloch and Pitts 1943) have been proposed in order to address this "distance" between the kind of digital computer that we use every day and the kind of distributed machinery that is actually found in real cognitive systems. Through connectionism a weaker notion of computation is surely achieved. While computations of early models heavily relied on symbolic-representational information processing, in artificial neural networks simply there's no use for the notion of symbol. Further on, ways has been found so that computation may survive the departure from the also influential notion of representation (Piccinini 2006) without having to fall into a contradiction. Even arguments advocating for an analogic nature of neural information processing didn't rule out the hypothesis that neural computation may well be a case of analogic computation (Trautteur 1999), or even a specific case of neural computation (Piccinini and Bahar 2013). Accepting the specificity of the neural substrate of cognition lead to accept that a computational system doesn't have to be computational at all level (Piccinini 2009). It can well be the case that computational processes may actually emerge from architecture that are only mechanical at the implementation level (Piccinini 2010a) and that, therefore, may even qualify ad dynamical.

From what we have seen, CH doesn't anymore take specific stance on what may qualify as computational. By doing so CH actually eludes a big deal of the original counter-arguments coming from dynamicist theories. Contemporary computationalism doesn't need to think at cognitive system as digital computers anymore and CH has become in response nothing more than "the view that intelligent behaviour is causally explained by computations performed by the agent's cognitive system (or brain)"(Piccinini 2009, p.1). Such a broad definition actually points out a very important fact: computation is indeed a flexible and therefore very resilient notion. Much of the explanatory constraints that DH considers unacceptable are derived from the strength of the definition of computation that is taken as fundamental. If in a cognitive system we can find processes that can be modelled in the form of an "effective procedure", then that is now enough to consider that cognitive system as computational. What really is important is not the general definition of computation, but the fact that the cognitively-relevant computation is actually found in the brain, which is cognitive per se. It's through this weakening of the definition of computation that complexity makes its way into CH. A more general notion of computation surely is able to accept a wider range of possible implementation while at the same time satisfying the need for a more faithful interpretation of cognitive and neurocognitive facts. However, there are also drawbacks. By accepting a notion of computation that kept on becoming weaker and weaker, the actual constraints coming from the implementation level gradually start to become more and more important. In order to consider computational cognitive system in complex way an account should be found for those constraints that come from the neural substrate inside the very notion of neural computation. At the same time, also environmental constraints have to be considered in order to reach a complete representation of what it means to be a real computational cognitive system. Complex cognitive systems, even

computational one, need to be considered then as intrinsically constrained by ecological and material boundaries. The complexity of the cognitive system may be a useful element of a good computational theory of the mind, but that doesn't mean that we must consider the fact that even if a computational system is complex enough to implement cognition that doesn't mean that cognition shouldn't be simple enough to be executed by a computational system.

4.4 Understanding complexity the computational way

At the start of the present section we have seen that contemporary computationalism claims that cognitive processes are computations (Cordeschi and Frixione 2007; Piccinini 2009), hence that a cognitive process always expresses one or more computable function (Massaro and Cowan 1993; Anderson 1990; Cummins 2000). We've also considered how applying complexity to CH leads, similarly to what happened for DH, to considering the cognitive system as a complex, more realistic, and architecturally constrained system. In the previous paragraph, we went in search for those constraints and we considered the option that cognitive systems should be considered as bounded. However, when talking in computational terms what we are looking for is a shift from plain computability to a version of it that admits the presence of certain performance boundaries.

Starting from computability it's well known that not every function is Turing-computable[45], but surely cannot pose too much of an obstacle. Computability checks are, for definition, inherently unbounded in a number of ways, and that is because they rely on an idealized computational model (the Turing Machine) that relies in infinite time and space resources (an infinite tape and infinite computational step). A function is computable in the

[45] See, for example, the halting problem (Davis 2004; Turing 1936).

classical sense if exists a finite effective procedure that halts after a finite number of steps, uses a finite number amount of storage and works for arbitrarily large set of inputs (Enderton 1977). This kind of solution has to be somewhat finite, but there's no a priori indication about how much memory or time does it needs in order to return an output.

As we have seen cognitive agents must be considered as complex cognitive systems and then we have to account for a number of ecological and material boundaries. We've already pointed out how real world decisions have to be taken in split seconds and actions are needed to start and stop at just the right moment. So, cognitive processes are bounded at one end by the task that is requested, but they are also limited at the other end by the characteristics of the computational machine (the brain) that executes the computations needed in order to perform that precise task. However, there are no precise estimates of the real computational power of the human brain, but as we have seen only a number of good reasons to believe in the finiteness of our brain capabilities. It seems then reasonable to focus on studying the computational resources requested by those problems that (we believe) our computational brain has to solve in order to express the large array of faculties and phenomena that we usually call mind.

If a cognitive system is computational and it's not trivially considered, then it has to perform not only the right effective procedure given the problem at hand, but it also needs to execute it in the right amount of time and using the right amount of memory in order for it to be cognitively feasible. This is a sense in which cognitive computations don't have only to be effective (computable) but also efficient and that has been widely explained in and introduced. The efficiency of computation is given by considering its tractability. As we have seen for this reason, the "tractable cognition thesis" have been proposed. This thesis claims that the mathematical theory of NP-completeness (Garey and Johnson 1979), and computational complexity as a

whole, may provide the theoretical constraints needed to account for the boundaries recognized by the bounded-brain hypothesis and the need for computationalism to account for complexity. Computational complexity is a theory about the hardness of problems, a theory that studies then the intrinsic complexity of computational task. It analyses how the computational resources (time and memory) required for solving a certain computational problem rise in function of a certain input size[46] and establish if certain cognitive problem should be considered as computationally tractable and, then, cognitively feasible. Considering sound this relation between computational tractability and cognitive plausibility it's not a trivial, because while it is true that cognitive systems seems to be somehow bounded at the neural level, there are also cases in which the human cognitive systems actually seem to solve computational non-tractable problems in the blink of an eye (van Rooij 2008, p. 954). So, in order to save the combination of the bounded brain hypothesis and tractable cognition there are two way of proceeding:

We can presume that the information utilized by the cognitive systems is way less that presumed and by doing so reduce our input size, but that may come at the expenses of psychological plausibility (less behavioural cues may imply, in theory, less propositional attitudes); Or else we can fiddle with the tractability threshold, usually fixed at the polynomial level, in order to admit complex, but also tractable, cognition in conjunction with psychological plausibility. Has we have introduced in chapter 2 the second alternative seems more promising and applies a particular account on computational complexity, called parameterized complexity (Rodney G Downey, Fellows, and Stege 1999), in order to devise a refinement of the tractable cognition

[46] The length of the string representing the input.

thesis, the FPT-Cognition thesis[47] (Van Rooij 2008). According to this view, the observation that some functions are non-tractable (non-polynomial) only in respect of some small aspects of the input (called the input parameter) may suggest that a number of cognitive capacities that are considered to be computationally non-tractable may be otherwise tractable in respect of a portion of the input size. This may also support the idea that it's not only the quantity of the input that constraints problem tractability, but also the quality of the input instances and, then, the theoretical and empirical criteria that define the available input contained in the information considered by the cognitive system.

Apart from the actual paradigm being used, what is important for us to note is that this notion of complexity poses a strong accent over the necessity of reaching a cognitively plausible explanation of cognition. Cognitive processes don't have to be only accurately represented in a psychologically plausible way, but they also have to pass a tractability check in order to be considered cognitively plausible. In this sense, computational complexity takes the shape of the "plausibility" notion individuated by Gervais and Weber (2012): "the probability that a model is correct in the assertions it makes regarding the parts and operations of the mechanism, i.e., that the model is correct as a description of the actual mechanism". This type of application tends toward an equilibrium between our naive experiences of cognition ad what we can reasonably think that our cognitive system may be capable of. In this equilibrium, however, it's the cognitive plausibility criterion that surely has the upper hand.

[47] That stands for Fixed Paramenter Cognition.

4.5 Two theories, two complexities and one goal

Ultimately what should be noted is that the notion of complexity interested both in computationalism and dynamicism shares a commune element. In these theories, a complex approach always involves appealing to model of cognition and of cognitive system that shows much more adherence to reality. DH realizes so by making a plea for a more adequate model of explanation that actually encompasses a wider range of possible cognitive phenomena. In dynamicism cognitive phenomena are what makes systems cognitive in the first place, so theories about cognition have to adapt to certain feature of the mind instead of constraining it. CH, on the other hand, reaches a complexity by focusing on a weaker definition for computation and computational systems. However, by doing so, computationalism also has to account for the rising importance of the features and characteristics of the actual neuronal implementation and so it has to introduce certain elements of complexity in its framework in order to demarcate what's relevant for cognition and everything else.

What both these theoretical hypothesis share is then the starting point Both of them are in search for a more satisfactory model of cognition. However, it is when complexity steps in that DH and CH part ways. In DH complexity is introduced with a clear emphasis toward psychological plausibility. Under this assumption the best model of cognition is the model that consent to better represent the range of cognitive phenomena usually ascribed to cognitive agents. Releasing cognitive science from the constraints of a certain computationalism makes possible to reach an explanation of cognition that's better because is more psychologically plausible. On the other hand, CH uses complexity the other way around. Computationalism had the progressive necessity of weakening its core notion of computation, but instead of embracing the primacy of psychological plausibility it has

implemented a notion of complexity with the explicit aim of restricting the range of what we can plausibly call cognitive. Linking computational tractability to cognitive plausibility provides a tool for accurately pinpointing a line of demarcation for distinguishing what's relevant for cognition from what, instead, is only a burden.

So, these two approaches at complexity are indeed very different, but it should be worth considering that even if the notions utilized by DH and CH are pointing in different directions they are indeed not contradictory in any sense. Psychological plausibility (or richness) and cognitive plausibility (or plausibility) are both good prerogative of good theories in cognitive science. Maximizing these two theoretical properties requires reaching a point of equilibrium that, on the other hand, would maximize the explanatory power of the particular explanation at hand. The contradiction that started the dispute can be considered genuine only if DH is evaluated in relation to its original target. However, such polemical target has gone through an evolution of its core notion so deep that it could be reasonable to argue that the eventual differences between a neural computational system and a cognitive dynamical system may null or negligibly small. In conclusion when we take into consideration the notion of complexity entailed by the dynamical and computational approaches what emerges is a concurrent relation between them that make them complementary in the journey toward a satisfactory theory of cognition.

Chapter 5

Another way of measuring complexity

In the latest two chapters, we have seen how complexity interacts with particular theoretical notions (simplicity being the case) and how it also influences and characterize entire frameworks (dynamicism and computationalism). Both the above cases make use in different degrees of computational complexity and then of a measure of complexity that we have learned to know better in chapter 2. Philosophy of mind is of course not devoid of other examples and that has been cleared at the beginning of the thesis. However, in the same occasion we have noted that different measures, or formal notion of complexity, are indeed interesting to look at because of the philosophical scope in which they find application. In this chapter, we will propose a slightly different take on the problem of assessing the role of computational complexity in cognitive science. Here we will first introduce one of the principal theoretical proposal that use a notion of complexity rigorous enough to be actually measured, Tononi's Conscious Complex Theory (CCT) (Tononi 2004b; Tononi, Edelman, and Sporns 1998; Tononi and Koch 2008; Tononi 2004a; Tononi 2010; Tononi 1998; Tononi 2012; Tononi 2003). A theory that takes the name of conscious complex theory (CCT) a that explicitly aimed at explaining what are the key properties that a system has to possess in order to express a precise feature, that of consciousness. In the second part of the chapter we will introduce computational complexity into the equation and see how it fares against CCT. In the conclusion of the section we will see these two complexity theories differ and what kind of characteristics arise from the confrontation.

5.1 Complexity according to Tononi

With the name of CCT we indicate the following thesis: human cerebral mechanisms express conscious experience because they perform a certain level of information integration. What is crucial of this thesis is the fact that it realizes a precise analogy between conscious experience and a notion of "information" that is not based on a special notion of content[48], but on a pretty specific form of information that have been "integrated" by multiple cerebral mechanisms that are organized in what is called by Tononi a “complex”. this way Tononi recognizes one of the principal features of consciousneess: its inherently multimodal but, at the same time, unitary nature[49]. However, beside the holistic element of consciousness just mentioned there is the necessity for a phenomenon, that is indeed minutely structured, to access this multitude of different states in order to make integration possible.

This is exactly the sense in which Tononi understands the relation between consciousness and "information". The internal distinction between the richness of numerous co-occurrent conscious states corresponds in a reduction of the uncertainty like the one that is obtained through a die roll. This theoreticla step leads Tononi into the adoption of Shannon’s information, a technical measure of information that we have already encountered in chapter one. As we have already said before SI performs a measure through a logarithminc encoding of surprise. Making an example will however be useful to catch the meaning of this: if we take the case of a coin toss, the event “obtaining head” will be determined by $log_2(2) = 1\ bit$ of information because only two possible outcomes are possible. The same can be applied for

[48] Like the one that may at the base of arguments like the knowledge of Mary (Jackson 1982) that about absent *qualia* (Chalmers 1996).

[49] The multimodal nature of conscious experience is confirmed by neurophysiologic phenomena like the refractory psychologic interval, that limits us to take only one conscious decision at a time in an interval of a few milliseconds (Pashler 1994), or the phenomenon of perceptive rivalry (Sengpiel 1997).

the tosing of a six faced dice, in this case we will have $log_2(6) = 2{,}58\ bit$. Even if taken, as it is, this way of measuring information may indeed seem a bit reductive when compared to the complexity of the human conscious brain, this notion serves as the foundation of Tononi's thesis. However, due to the well known phenomenic aspect of consciiousness even an unfatomable amount of information would not be enough to explain consciousness. The solution of Tononi is to look beyond information in itself and instead go in search for those modalities that permit the integration of different informational states into complexes. This is the reaosn why the mechanism of integration is at the heart of Tononi proposal of considering the capacity of integrating information necessary for consciousness. The evidences that ground this intuition come from a range of experimental findings in neurophisiology. The transition between automatic unconscious tasks and conscious ones is accompanied by a clear change in the neuronal activation patterns. In the case of a new and unencountered before tasks, the neuronal activity show a widely distributed area of activation. The same task, once learned and automatized, will show not only a different pattern of activation, but also a different and more localized modality of it (Petersen et al. 1998).

This phenomenon of neuronal segregation contextuale with automatic tasks seems to be provide and advantage in speed and execution economy, while paying a price in terms of context sensitivity and flexibility (Baars 1988). Furthemore, Tononi recognizes that all conscious tasks are also accompanied by activation of functional dynamical clusters that emerge in precise time intervals and are composed of neuronal units with a high internal coherency. These unit, or complexes, show a high propention at constituiting nervous links with other cluster as well as single neurons.

It is such contextuality between conscious tasks and activation patterns that allows Tononi to suggest that such patterns, along with their functional characteristics, are to be considered the condition for consciousness

alltogether. Individuating the relevant neural correlates is a thing, explaining the reason why they are so is another and in Tononi 's theory such role is filled by the above mentioned information integration. Tononi's arguments goes like this: if conscious experience is correlated with certain neuronal clusters, and if the property of such clusters can be individuated in information integration, then conscious experience will be strongly correlated with the capacity of a system to integrate information. This approach has indeed two main advatages: it practically allows to measure consciousness; it allows to evaluate the neuronal correlates and according to their presumed capacity of integrating information, hence their capacity of gathering themeselves into what Tononi calls functional cluster. In order to betterunderstand what are the implications behind this choice we have to take a deeper look into che technical notion that Tononi uses in its thesis. As we have previously mentioned Shannon's information, or entropy, is a measure of uncertainty and can also be a measure as a measure of the states variability in a system. The following is the formal way of defining it:[50]

1. Given a system X composed by a set of elements $\{x_i\}$ so that X can assume $m = 1 \dots M$ discrete states;
2. Assume that any of these states is associated to a probability value p_m, so that the sum is equal to 1;
3. For X, entropy can be defined as follow:

$$H(X) = -\sum_{m=1}^{M} p_m \log_2(p_m)$$

[50] See Cover and Thomas (2006) for the full demonstration.

Being entropy a measure of uncertainty, its magnitude will be propotional to the number of equiprobable states that a system can take. The value of entropy will be instead equal to zero if the system X takes only one state with $p = 1$, that is certainty. Once that the entropy value has been considerates it is possible to investigate what really is a functional cluster according to Tononi and how to recognize it. At least intuitively it is possible to define a functional unit as a subsystem, a single indipendent (according to certain properties) part the main system. The same holds for cerebral areas: they are components of the brain that can be distinguished for anatomic and functional reasons. These functional reasons will make so that clusters and functional units will be statistically salient inside the main system. They will then realize a sort of "reduction of uncertainty" that can be captured through Shannon's entropy. This is exacly what Tononi's suggests. By considering the subsets of a system, in this case the human brain, it will be possible to register the statistical dependedancies between the subsests themeselves and between each subset and the whole system. Again, this statistical dependency will be captured by a tool coming from information theory: mutual information (MI). MI "is the reduction in the uncertainty of one random variable due to the knowledge of the other", (Cover and Thomas 2006, p.19). We can then formally define MI in the following way:Possiamo quindi rappresentare MI nel seguente modo:

$$MI\left(X_j^k; X - X_j^k\right) = H\left(X_j^k\right) + H\left(X - X_j^k\right) - H(X)$$

Putted in plain english: the mutual infromation between a (neuronal) subset and the whole systems is given by the sum of the entropies of the subset and the system without the substet minus the entropy of the whole system. This way the actual contribution of the considered subset will be confronted with the original entropy of the system. It is important to note

that, contrary to what happens for entropy, MI is a positive and simmetric measure. It will assume the value 0 in case of statistical independency and every other positive value in case of statistical dependency.

MI measures then the degree of relation that a subset has with its system. However, another component of a functional cluster to consider is also its internal coherency and complexity. It is reasonable to suppose that a "valuable enough" subset will display a high internal coherencly and then be also characterized by a high degree of causal dependency between the single elements that compose the subsystem. Following the same intuition that grounds the adoption of of MI, Tononi proposes again to measure such dependency by verifying the loss of total entropy in the subset itself.[53] This loss of entropy is captured by Integration I(X) and is defined as the differences between the sum of the individual entropies of the components of a system and that of the entire system:

$$I\left(X_j^k\right) = \sum H(x_i) - H(X_j^k)$$

I(X) will be zero if the component of the systems are statistically independent, vice versa instead if statistical dependency is present. Together MI and I account from one of the definitory traits of what Tononi calls a "complex". Being that the capacity of taking the form of a coherent and cohesive functional unit that is the protagonist of a signitificant relation with the system in which the subset resides.

If we translate the results above into a more cognitivistic vocabulary: A functional cluster can by consequence be defined as a subset of cerebral regions that displays a certain degree of integration I that is higher than the

[53] Here treated as a whole independent system in itself.

MI between the subset itself ant the rest of the brain. Both these measures find a synthesis into the following *Cluster Index* (CI):

$$CI\left(X_j^k\right) = \frac{I(X_j^k)}{MI(X_j^k; X - X_j^k)}$$

From the above formula emerges that CI will rise proportionally with the value of I, representing the internal coherency and cohesion of the complex. Instead, the value of CI will be inversely proportional to MI, representing the fact that the relations inside the subset need to be stronger than those with the whole system. If CI is equal to 1 this indicates that the subset's internal statistical dependence I is equal to MI. This means that the functional cluster is practically indistinguishable from the system that contains it. In this situation, no complex will arise inside the brain. CI is however not an assertive measure, but provides an indicative tool to individuate those neuronal subsets that effectively qualify as candidate complexes. Once a number of candidates are selected, it will be possible to further select those complexes that display the right feature to be considered genuine functional cluster so that their intrinsic complexity can be finally evaluated. It is in this final evaluation that we reach the full-fledged measure of complexity in which we were interested from the start. Tononi's proposes that to find the right synthesis between the information expressed by a system[54], the mutual information MI and the integration I, we have to address a measure of Neuronal Complexity (CN). The aim of this measure is "to estimate the average integration for subsets of the neural system of increasing size; that is, at multiple spatial scales" (Tononi, Edelman, and Sporns 1998, p. 476). To quantify then the difference that a cluster makes in a neuronal

[54] Here intended as a population of events all having a certain probability value.

system we should not consider a single subset, but all the possible combinations of subsets. This leads to the following formula:

$$C_N(X) = {}^1\!/_2 \sum_{k=1}^{n} \langle MI(X_j^k; X - X_j^k) \rangle$$

The measure of complexity here presented is function of the average MI between every subset and the rest of the system. C is high only if there are subsets in the system, if these qualify as functional clusters and if between these subsets are present enough statistical dependencies to imply the existence of a relevant and genuine functional relation with the system. The complexity C does make the case, in Tononi's intention, for the information integrated by the complexes. This emerges from the following quote:

> *"Complexity is mathematically equivalent to the average information exchanged between subsets of a neural system and the rest of the syste*m, *summed all over subset sizes. Thus, complexity provides a measure for the amount of information that is integrated within a neural system."(Tononi 1998)*

Tononi's claim can then be now rephrased as the following: the human brain is conscious because it is an ensemble of dynamical and salient functional clusters that integrate states (information) into a unified scenario (consciousness). Beside indicating a measure of complexity (Tononi and Balduzzi 2009) also proposed and index, similar to CI, that expresses the value of integration that a certain functional unity can reach. The idea behind this further proposal is indeed linked to experimental practice and is based on evaluating how a subset reacts to a perturbation of some sort. The reaction that is obtained by consequence will provide an indication about the characteristics of the chosen subset. This intuition can be again translated into

a formal definition that characterize the integration index by generalizing the relative entropy (RE) on the power set S:

$$\Phi(x_i) = H\left[p(X_0 \to x_1) \parallel \prod_{M^k \in \mu^{min}} p(M_0^k \to \mu_1^k)\right]$$

Again, we will provide a translation of the above formula so that its key elements are easier to handle. Tononi, in order to capture the sort of additional information that a cluster contributes to, proposes to evaluate the ER of the entire system in respect of the considered cluster. This cluster, however, does not have to a specific one. The whole system is decomposed into its littlest informative partition (μ^{min}). The integration value of a cognitive system, hence its complexity, will be as high as that of its μ^{min}. If such system can be partitioned into informationally insignificant and independent elements, then it will express low value of integration and complexity. Φ can be considered as a measure of the causal relevance of the μ^{min} , and then of the smallest functional cluster present in the system.

If we go back to consciousness and integration again, we can suppose that it would be possible to track in the human brain architecture hints of this capacity of integrating different states. Tononi's theory seems to be consistent with a certain number of theoretical conjecture and also with some experimental evidence. Thalamocortical regions, for example, are often been proposed as the seat of consciousness (Plum 1991) and this seems to be corroborated when such regions are considered under Tononi's assumption. These regions show a value of integration that is, for example, much higher than that the cerebellum, which has a higher neuronal count.

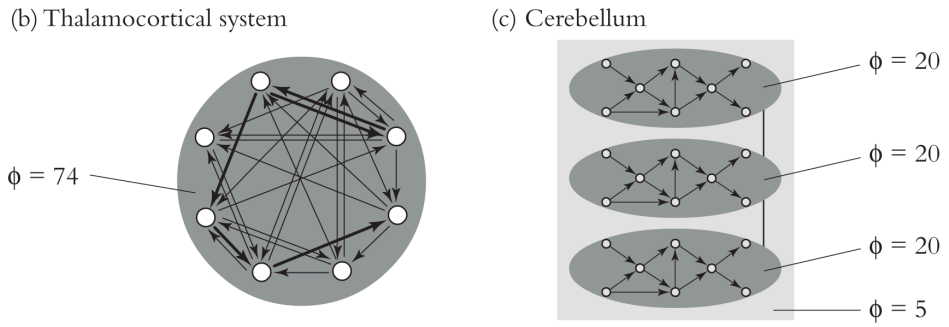


Figure 1 Tononi (2004) An information integration theory of consciousness

In the above figure we can see how the thalamocortical system, exemplified here through a model (Tononi 2004a), possesses a integration value much higher than that of the cerebellum. In this case the correlation between the neuronal activity and the relative conscious states is determined by the fact that these regions behaves as complexes and are then able to integrate information accordingly. More evidence favouring Tononi's account come from the consideration of cases of intense stimulation of wide cortical area normally associated with epileptic seizures. In these cases, even if a wide activation pattern is present, no functional cluster is able to emerge and then also no consciousness. Evidence may also come from the split-brain phenomenon, that may provide an example of the dynamic and flexible nature of functional cluster (Sperry 1976; Gazzaniga 2005). As already mentioned, Tononi's framework also has various theoretical allies. The *Global Workspace* (Baars 2005; Baars 1988) theory, together with the blackboard metaphor of Dennett (1991) and the *Global Neural Workspace* (Dehaene and Changeux 2011) are all accounts that support the idea of consciousness as a merger of states and provider of unified access to the external world.

Now that we have analysed Tononi's proposal in detail we can shift our focus towards the way it can be collocated in the complexity landscape. Again, we will adopt the confrontation method and see how Tononi's

complexity behaves when it is put against another formal measure of complexity that can be cognitively applied, computational complexity.

5.2 Where does Tononi's complexity stands

In the first chapter, we have seen how complexity often takes the form of a synthetic measure that accounts for those features that are considered relevant for a system. Those features may be relevant for a particular cognitive capacity (mindreading) or for a more holistic property of the systems (being cognitive or conscious). There are also numerous formal ways to express complexity, however we have also seen briefly that such measures even though formal may not be theoretically transparent. Intuitively, they seem to behave more like tools and like tools are more suited towards certain applications. According to this idea it would be plausible to also evaluate Tononi complexity not only in a direct fashion, as we have done in the paragraph above, but also by considering the implications are behind it. Here we will argue that Tononi's approach to complexity, besides being a formal and rigorous approach to complexity, has three characteristics that ultimately collocates it in a precise epistemological spot.

The first aspect that we will consider comes from Shannon's Information (SI) and then from the technical tools that Tononi employed to capture his idea of complexity. SI is really a notion of information, but as we have seen in chapter one, can be conceived as a notion that performs a measure of complexity. SI is influenced on the one hand by its reference unit, in this case the *bit*, and on the other hand the phenomenon that it has been designed to measure. These two elements make so that only statistical dependencies can be meaningfully and reliably treated by SI and, therefore, by all of its derivatives. Furthermore, the unit of measure itself if the consequence of a precise encoding and expresses a magnitude in function of the length of it.

That qualifies SI as a measure of complexity that favour notions such as description more than difficulty, or effort. Accordingly, Tononi's complexity will plausibly share the same conceptual preference toward description. This appears to be consistent with the fact that this type of complexity seems particularly concerned with the organizational features of systems, the way in which systems last can be partitioned and the relations that are present in them. This would at least indicate that Tononi complexity could be indeed explicatively partial and then unbalanced toward certain problems, phenomena and notions. Taken in isolation this aspect would not be particularly problematic. However, when the particular application of Tononi's account is taken into the equation it can bring to some flaws.

This brings us to the second aspect of Tononi's complexity that we consider interesting: its link with consciousness. The complexity notion that he proposes is in fact openly domain-specific and tailored to the one phenomenon that is consciousness. This has two main consequence: the first is that if kind of complexity measured is unbalanced toward description it will remain unbalanced also in regard to the specific phenomena that it is addressing. The second consequence comes instead from the way in which complexity is employed. Tononi is not concerned with finding a definition of complexity per se but only in relation of complexity. He uses a measure of complexity as a condition for consciousness and builds on it the following argument: if a system express a certain value of complexity (equal or superior to the human one) then it will be conscious. For starters, this type of inference is inherently biased by the fact that the nervous system from which the neurobiological evidence is taken is the human one. If we add this to the above mentioned descriptive bias of the chosen measure of complexity, we have that not only the inference toward consciousness will be biased, but it will also provide a partial account on the nature of consciousness itself. All these points sum up to the fact that while in the intentions of Tononi his

complexity seems aimed toward individuating a sufficient condition for consciousness, what it really constitutes is only a necessary condition for it. This means that, while the architectural features individuated by this type of complexity are still crucial, they can be associated with other requirement that explain consciousness from another perspective.

To better collocate the role that Tononi's complexity in cognitive science plays we can now take into account David Marr's widely cited level of analysis[59] (Marr 1982) and our previous considerations about the status of computational complexity. Marr's describes three levels at which an "information processing device" can be analysed:

1. Computational level:

   Here the computational problem is stated through the definition of its presumed inputs and the desideride output. No indication is given here about how the computation is actually performed;

2. Algorithmic level:

   Here a representation or algorithm of the computational problem stated in the first level is chosen and then defined. No indication is given about the actual concrete realization of such algorithm.

3. Implementation level:

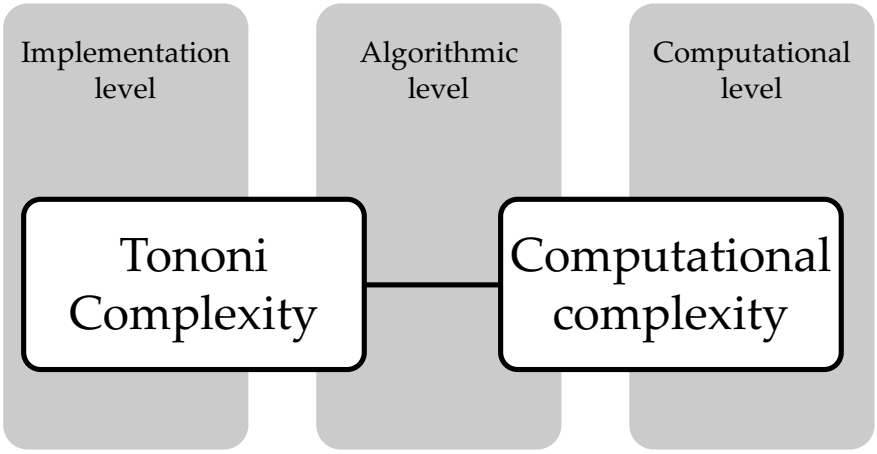
   Here the actual physical implementation of the algorithm that solves the original computational problem is defined.

What we propose is a slightly different from the usual utilization of the above the above hierarchy. By considering the requirements of each level in isolation

[59] While Tononi's proposal never explicitly mentions its "membership" to the computationalism framework, it also never rules out the validity of it.

we can individuate the characteristics that a hypothetical notion of complexity would need in order to be there applied. Such a methodology can then be used to extract a reference system that makes possible to hypothesize different explanatory roles for each of the complexity notions here considered. Computational complexity provides here a major hint on how to proceed. In chapter one we indicated that computational complexity is not a general notion of complexity. For starters, it is limited to the acceptance of the computationalist framework, and then of the possibility to have a computational explanation of cognitive phenomena, consciousness included. Furthermore, computational complexity cannot be general even in these computational explanation, but its application is limited to the computational level of Marr. That is given by the fact that computational complexity is a theory about the hardness of computational (computable) problems and not of specific algorithms or, even worse, physical computational systems. The taxonomy of computational complexity classes can only (or at least conventionally) be used at the computational level of Marr. From this follows of course the original need for the tractable cognition thesis to always have a translation of cognitive capacities into computable functions and therefore computational problems. If we apply now the same methodology to Tononi's account we have the following: it utilizes a measure of complexity that favour description over execution, it is particularly concerned with the architecture of systems more than in their performance and it qualifies as domain-specific measure that is related to a single cognitive capacity or phenomenon, namely consciousness. The first two key aspects of Tononi's proposal provide an indication of where to put its complexity inside Marr's three levels, the implementation level. Some may argue that Tononi's complexity doesn't provide any information about the actual physical components of a conscious systems, but only of the organization requirements that a system should have in order to sustain consciousness. The implementation level has been indeed

thought as the seat for details about the implementation of computational systems, however we have to remember that complexities are not explanations. They are instead notions that are built upon explanations, model and theories, and then by consequence they will be an abstraction from the source theoretical substrate. The following diagram show how Tononi's complexity notion fits inside Marr's hierarchy:



Tononi complexity is then a good tool for characterizing in a synthetical and practical way what are the architectural proprieties that a certain implementation needs to possess in order to support consciousness. However, there is also another last aspect to take into consideration before ending the chapter. While the above characterization of Tononi's complexity highlights its collocation in a cognitivistic and computational explanatory framework, it also accounts for its partiality and inherently biased nature. We have already hinted at the fact that even limiting our scope to consciousness this type of architectural complexity has the chance of only providing necessary conditions for it. This is only accentuated by the collocation inside the implementation level. A level of analysis that is inherently underspecified in respect of the original computational problem that stands at the other end of the hierarchy. No indication is given through Tononi's complexity about

the effort that a cognitive system should take for performing conscious tasks, or about the resources that are needed for it and, ultimately if integration is actually a problem whose difficulty makes it plausible. If the complexity of consciousness should be evaluated completely and thoroughly these kinds of questions should be at least addressed. I see the application of the tractable cognition as a viable option here and, even if no formal model of "cognitive integration" in the sense of Tononi is available we can at least informally define the issue through the following steps:

1. Integration is the base process that Tononi's complexity captures;
2. If the capacity of integrating multiple states is necessary for consciousness, then the complexity this capacity should be not only architecturally measured, but also translated into a computational problem;
3. The definition of integration problem, once stated, should be considered under the tractable cognition requirements. Otherwise theory revision should be considered.

Ways of exemplifying the process of integration that Tononi wants to explain and ground in complexity can be found, for examples, in technical applications like *referential integrity* utilized for merging data received from heterogeneous sources[73], for which computational complexity results already exist (Chomicki and Marcinkowski 2005). The integration process is well defined on the anatomic and neurophysiologic part, but finding an accepted model of its neuronal implementation will take time. However, it is possible to partially find an answer though the evaluation of the various states that consciousness is supposed to integrate. What emerges from experiments and

[73] Systems that realize an aggregation under other unique coding of data from sources characterized by different originating forms of encoding (Liggins, Hall, and Llinas 2008).

from folk psychology seems to indicate conscious experience is indeed composed by numerous cognitive modalities. Among them we find of course phenomena like visual search, inferential reasoning, learning and also language production and comprehension (Dehaene and Changeux 2011). Various computational models proposed for the above-mentioned phenomena and capacities are however computationally intractable (Van Rooij 2008; Wareham et al. 2011) and since computational intractability travels from special cases to more general one, we may have good reasons to believe that the process proposes by Tononi may be intractable as well. For the time being the only cleat indication about the computational feasibility of Tononi's proposal come from Maguire et al. (2014). Here the integration process is that's crucial for the measure of complexity that Tononi's uses is equated to "complete lossless integration" and evaluated as such. The authors conclude that, if the above equation holds, "complete lossless integration requires non-computable functions", meaning that one of the crucial premises of Tononi's complexity may be computationally impossible.

For concluding, in this chapter we have evaluated a formal notion of complexity together with an application of it that is both philosophically relevant and cognitively relevant. The considerations that we have advanced and the results here considered corroborate one of the hypothesis that we have advanced in the first two chapters of the thesis. Complexity is indeed sensitive to its application and the investigation of it deeply enhanced by comparing different notions, in this case Tononi's complexity and computational complexity, between them and in respect of more general explanatory framework (Marr's three level). Applying this methodology to Tononi complexity made possible to assess its core aspects and characteristic, but crucially provided also additional hints on how computational complexity behaves in relation with other formal measure of complexity that can also be applied to cognitive phenomena and capacities.

Chapter 6

Making mindreading tractable

In this chapter, we will propose how the TCC can be applied to a specific cognitive capacity, namely mindreading, and see some of the consequences that the application of the thesis has not on the actual philosophical and psychological definition of it. Mindreading can be broadly defined as that cognitive capacity that allows an agent to understand, predict and explain his behaviour and that of other agents as well. This capacity also takes a number of other names and labels depending on the theoretical framework of reference. "Theory of mind" or "folk psychology" are only the most famous and adopted examples. Here however we will adopt "mindreading to refer to the range of philosophical concepts that have been introduced into the study of social cognition, but principally to indicate the actual cognitive capacity that takes external behaviours as input and outputs an interpretation for them. However, while a coarse functional definition of mindreading may be easy to state, determining how mindreading is actually able to perform such function is indeed tricky.

A lot of different accounts have been proposed in the literature: ranging from the theory-theory account (Carruthers and Smith 1996), to the simulative account (Goldman 2006a) and, more recently, the enactive account (Hutto 2015). It is important to remark that all of these different theoretical proposals do not differ in the actual role of mindreading and the phenomena that are correlated to it. The main theoretical culprit is instead the way in which mindreading works and that is reflected by the fact that the main controversy is indeed on the inferential nature of mindreading. That is on the fact that

mindreading is indeed based on a form of inference where external behaviours are premises and interpretations are the conclusions, in the middle stands the background (social) knowledge needed for the inferential step. The present solution is obviously advocated by theory-theory (TT) accounts of mindreading, but is also pretty popular in computational and bayesian modelling of social cognition capacities (C. L. Baker, Saxe, and Tenenbaum 2009b; C. L. Baker, Tenenbaum, and Saxe 1995; C. Baker and Tenenbaum 2006). Simulative (ST) and enactive accounts (*a fortiori*) both negate that and instead propose different ways in which mindreading should be intended: the first account propose that instead that on inference mindreading is based on mentalization and on the capacity of simulating actions and behaviour of our conspecifics; enactive accounts negate directly mindreading by proposing that simple action coordination, and maybe more, only needs automatic and non-representational behaviours. Both of the criticisms here presented are based on empirical evidence, but also on a plausibility argument. Inferential mindreading should be considered false principally because it is not cognitively plausible and, especially, too greedy on resources (Gordon 2009).

In the previous chapters, we have seen how computational complexity is used by the TCC as a measure of plausibility for cognitive theories. Since the debate on mindreading is indeed centred around on the plausibility of it, seems plausible to try an application of the TCC in order to see how the mindreading capacity does react to various types of restrictions and how the philosophical claims on mindreading can be reconceived once that computational efficiency is taken into account. To do so we will propose the following applications: in the first section, we will consider the specification of the intentional content in the attribution of psychological states and argue that such a specification cannot be complete without strong indications of it being also computationally intractable; in the second and last part of the

chapter, we will see how Mindshaping, a theory that propose that the cultural and social context makes mindreading tractable, can be operationalized and pass the muster of formal computational complexity analysis.

6.1 Mindreading, tractability and intentional content

The following popular quote from Friedrich Nietzsche by itself may work as a sort of *leit-motiv* of the 20th century: "All that exists consists of interpretations"(Nietzsche 2011). Positivism instead refuses to attribute such an importance to subjectivism and stops at phenomena stating: “There are only facts and nothing more.” In opposition to this way of thinking I will imaginarily take the side of Nietzsche and say that facts are precisely what is lacking in human experience, all that exists has to consist of some sort of interpretations. We cannot establish any fact “in itself” and, in this time and place, it may even be nonsensical to aspire at the complete objectivity of facts. To the extent that knowledge has any sense at all, the world is knowable: but it may be interpreted differently, there is then not only one sense behind every phenomenon, but hundreds of senses.

In the philosophical debate a number of objections have been proposed against this subjectivist way of reasoning. The supporters of "new realism" theories consider assuming that there is no direct access to the world simply a mistake. They refuse to accept that the data of experience exist only within conceptual scheme, and that in the end knowledge, much like human experience, may be nothing more than a bundles of opinions imbued with culture, language, symbolic forms, signs, and social conventions (Ferraris 2012; De Caro and Ferraris 2012). It is not true that the only epistemic space open for business is the market for different interpretations. On the contrary, the world is able to do a real friction towards our conceptual frameworks. It

draws, in fact, the field of those interpretations that are interpretations *of something*.

There is a wide debate about hermeneutics, cultural studies, and the new realism. However, whatever you think about it, it is noteworthy that the wind of the interpretation has finally come to the field of cognitive science. In the late seventies and the eighties of the last century, several scholars have emphasized the idea that social relations are mediated by *psychological* interpretations. Among all the phenomena that seemed expecting to become the object of a process of interpretation, the race was finally won by the mind. From the objects of interpretation, such as texts and symbolic forms, the enthusiasm for the interpretation has come to engage directly with the individual who produces it, or, more precisely, his or her mind. It was virtually inevitable: if the interpretation refers to signs and symbolic forms, it does so by virtue of a mind. The signs, in fact, as Charles Sanders Peirce states, are always representations and these latter, in their turn, are representations only if they are considered to be part of the realm of the mind. In his words: "A sign is something which stands for another thing *to a mind*" (Peirce, 1873 - MS 380, my italics).

Mental interpretation (or mindreading), however, is not an "all or nothing" phenomenon. It comes in degrees and has components. In the current literature, it is common to draw the distinction between two levels of mindreading (Goldman 2006b; Coricelli 2005). The first level is represented by a low-level simulation concerning the understanding of the aim of an action, and the other consists of a high-level simulation taking place in cognitive processes such as the taking of a different point of view from one's own, and the so-called "counterfactual imagination". The neurophysiological basis of low level of simulation mainly consists in the mirror system and in the cerebellum (Gallese et al. 1996; Rizzolatti and Sinigaglia 2008; Ito 2012; Perconti 2015).

High level simulation is an activity of projection which, to take place, must have an inner space in which to be based and from which to operate. Reflexive reasoning, i.e., the linguistic side of self-consciousness, is the inner space from which high level simulation proceeds in its attribution of intentions, and in its behavioural predictions (Perconti 2008). The idea that reflexive reasoning could work as a base for high level simulation, and as inner space for behavioural predictions, is also a way of making possible for simulationism to tackle the problem of the behavioural prediction in competitive situations. To explain behavioural prediction in competitive situations, in playing games, and in erotic stimulation, the simulationist's approach must be able to distinguish between what I would do in counterfactual circumstances, and what, instead, I would expect that the individual I am simulating would do.

Differently from the high level of simulation, the lowest one doesn't need any intentional attribution. Understanding the goals of the actions is an automatic process which relies on hard-wired mechanisms in the brain. It is not a conscious and voluntary activity. This means that we understand other people's actions, intentions, and emotions even without any conscious mental representation. Or, in epistemological terms, to understand an action goal and the others' emotions we can do without the intentional account on mental representation with its typical vocabulary made up of "aboutness" and "mental content". Moreover, in the low level of mindreading we haven't a "propositional content" and, then, we might do not consider the influence of language in its characterization. On the contrary, the high level of mindreading implies the whole theoretical framework of propositional intentionality. This means, among other things, that it is impossible to attribute a mental state, without attributing also a propositional *content*.

6.1.1 Specifying the intentional content

The mischaracterization of the intentional content will be the main topic of the present paragraph. In order to attribute a mental state to another individual it is necessary to be able to specify its content. Both the theoretical and practical aspects surrounding this supposed specification is however a highly controversial matter. To appreciate the theoretical problem of how to characterize the intentional content in mindreading it can be useful to take into consideration how the attribution of representations to other animals works. According to Jacob Beck (2013), the state of the art on animal cognition depends on the conjunction of the two following theses:

> *"Realism: Animals have causally efficacious cognitive representations with determinate contents.*
>
> *Indeterminacy: We are currently unable to provide precise linguistic characterizations of the contents of animal's cognitive representations" (Beck, 2013)*

Realism is supported by decades of empirical research in the field of animal psychology and cognitive ethology. After the collapse of the typical skepticism of 20th century on mental representation, "most animal researchers now accept that animal cognition involves operations over causally efficacious representations with intentional content - representations that *characterize the world as being a certain way*" (Beck, 2013, p. 520; my italics). The hard problem of realism regarding animal representations consists in this characterization. It is widely - but most of the times implicitly - accepted that it is matter of a *complete* specification. According to the mainstream perspective, to be able to specify a given mental content requires the ability to explicate all the properties included in that content. Without fulfilling this

requirement, it should be impossible to distinguish a given intentional content from another one.

It is precisely because we are unable to provide complete characterizations of the intentional content of animal's representations (indeterminacy thesis) that one can suppose that other animals don't think at all. For example, this is the claim of Donald Davidson's in *Thought and Talk* (1975). Since we are unable to completely specify the other animals' intentional content, we come to the conclusion that they don't possess have thoughts. The problem is with the role that language plays in the specification of the intentional content. This specification is, in fact, conceived as nothing more than the possibility of translating a certain animal signal in a corresponding linguistic expression. And this seems actually impossible since there is not thesaurus available for non-human languages.

For this reason, Davidson (1975), as well as Daniel Dennett (1989) and Dale Jamieson (2009), believes that we have not to take too seriously the practice of attributing thoughts to animals. These attributions are useful explanations of animal behaviour, but they are not literally true and not necessarily reliable. In Beck's words (2012):

> *"It is not literally true that scrub jays have beliefs about the food they cache, or that chimpanzees believe that bare branches facilitate termite fishing. Such explanations are to be taken no more seriously than explanations of a thermostat's behaviour in terms of its 'desire' to keep the room at 72 degrees Fahrenheit, and its 'belief' that the room has deviated from that temperature."*

It is a kind of interpretationism, i.e., the account on mental reality which is committed to the convenience of the intentional stance, but not to the ontology of the realm of the mind. We can attribute beliefs to other animals purely instrumentally, in an ontologically frugal way. However, if we take

seriously the ontological commitment in the animal cognition, we have to admit that the attribution of beliefs to other animals requires that we can accurately describe their intentional content. But, if this is the case, as Steven Stich (1979) maintains, we have to conclude that animals don't have beliefs at all.

According to Beck (2013 p. 525), we should assume that "sentences have determinate contents, and that those contents are *at least* as fine grained as the sets of possible worlds circumscribed by their truth conditions". But, how should we consider the expression "at least" in the above proposition? The "at least" clause has more profound effects than one might expect. In any communication exchange the specification of the intentional content occurs in a partial way. In order to understand the intentions of other people it is not necessary a full understanding of their intentions. This may happen mainly due to time constraints and cognitive economy. Mindreading is a defining capacity of human cognition. It is believed to be always at work in our lives and also to be the basis of the human proficiency at social interaction and coordination. However, we should not fall to the temptation[77] of considering mindreading a special and unbounded capacity that is free from limits. As we have seen in chapter two the human cognitive system as well as the tasks that it can perform both have boundaries. So, also mindreading should be thought as a tractable cognitive capacity, which cannot use more cognitive resources than those that are actually available. It is just this tendency inherent in the human mind that prompted the low level of mindreading to become a cognitive process which is largely automatic and involuntary.

The above considerations lead us to think that we need a theory of mind that makes explicit the requirement of partial specification of the intentional

[77] Even the philosophical one.

content. In what follows we will try to show how this requirement may be grounded into computational tractability.

6.1.2 From mindreading to tractable mindreading

Having introduced cognitive tractability in chapter two we will skip the introduction of the tractable cognition framework and delve directly into the matter at hand. Mindreading as a cognitive capacity involves the ability of predicting and explaining other agents behaviour and the ability to discriminate agents in the environment and attribute mental states to them (S. Stich and Ravenscroft 1994). In order realize such a task a "mindreader" is supposed to use observable behaviour of other cognitive agents and attribute propositional contents in virtue of that alone. However, such a task seems to be utterly impossible to tackle because of the known problem of "holism" that mindreading has to face. According to that “any observable behaviour is compatible with any finite set of propositional attitudes, accurate propositional attitude attribution that is timely enough to make a difference to behavioural prediction in dynamic, quotidian contexts appears to be computationally non-tractable” (Zawidzki, 2013, p. 134). This version of the holism problem is indeed linked to various other philosophical problems that point toward the critical nature of many to many relation between sets and the capacity of making reliable inference from them. Hume’s induction problem states that “instances of which we have had no experience resemble those of which we have had experience” (Hume 2012). Also the frame problem (Dennett 1978) is in itself a statement about the impossibility of finding the right kind of restriction on a domain, so it may as well be considered as a reason for the supposed intractability of mindreading. Going back to holism, since it's logically possible for an indefinite number of behavioural instances to be mapped to an equally indefinite number of propositional contents possessed by an individual and *vice versa*, the search

space for a mindreading cognitive capacity will always be overly demanding and then computationally non-tractable. Furthermore, intuitively we may think that by adding a completeness requirement on intentional knowledge for propositional contents what we face is a situation in which the already overwhelmed mind-reader is burdened with another demanding task. Ironically however this may not be the case because if mindreading is inherently flawed by holism it also means that reaching a complete specification of the intentional content would be theoretically impossible and, a fortiori, also computationally intractable. In light of this, we will consider these two "flaws" of philosophical mindreading separately.

Nevertheless, human mindreading seems to pretty solid and widely used in a huge number of social interaction. This is surprising especially considering the strict time constraints in which we usually operate such a task. So, if we rule out the hypothesis about the special and unbounded nature of mindreading and accept that cognition has boundaries fixed by computational tractability, we have to take into consideration ways of tackling the mindreading problem without assuming an unreasonably powerful theory of mind. The problem of mindreading has to be reconsidered in at least two ways: by refusing the completeness requirement on propositional content and then by defusing holism.

As Zawidzki (2013, p. 161) notes, the problem of holism seems to be an almost exclusively philosopher's problem. Logical possibility surely opens up the chance for indefinite behaviours and mapping to propositional contents, but there are reasons to believe that "Human cognitive heterogeneity is easily overstated" (Zawidzki, 2013, p. 162) and that successful mindreading depends on the capacity of humans to overcome variability between individuals by relying on common social ground. In this sense the most of mindreading tractability is made out of social "mindshaping" processes that consists in a sort of sociocognitive priming over the most probable

propositional attitude given the social interaction at hand and the accessible behavioural cues. What this priming mechanism does is heavily rely on socio-cognitive and situational constraints in order to reduce the search space of a hypothetical mindreader, integrating at the same time the type of psychological constraints advocated by the fast and frugal heuristic hypothesis (Carruthers 2006; Goldman 2006b). What should be noted is that even if this kind of evaluation surely try to reduce the possible inputs only to a small relevant set, even if Zawidzky is right about mind-shaping, there are doubts about the feasibility of fast and frugal heuristics (Van Rooij 2008; Van Rooij, Wright, and Wareham 2010). Moreover, there may also be concerns about the tractability of mind-shaping capacities themselves, since it's reasonable to think that they may rely on some of the cognitive capacities of mindreading, such as, for example, mental states attribution from behavioural cues.

However, even if thanks to the proposals above holism cannot be considered a real threat to tractable mindreading anymore, a complete knowledge requirement of propositional contents it's still a problem for tractable mindreading. If such a requirement has to be satisfied, even a small input set my not suffice since the information about the propositional content should already be completely available in the input set, making the inference from observable behaviour to propositional attitudes demanding and, more importantly, completely useless. Even considering a complete homogeneity between individuals it's quite difficult to accept the idea of a complete accessibility of information about propositional contents. So, there are at least two way of proceeding from that: reducing further the presumed input size, but that may come at the expenses of psychological plausibility (less behavioural cues may imply less propositional attitudes, actually reducing the complexity of human interaction), or relaxing the already mentioned polynomial requirements for tractability in order to admit complex, but also

tractable, cognition in conjunction with psychological plausibility. The second alternative seems more promising and applies a particular account on computational complexity, called parameterized complexity (Rodney G Downey, Fellows, and Stege 1999), in order to devise a refinement of the tractable cognition thesis, the FPT-Cognition thesis [78] (Van Rooij 2008). According to this view, the observation that some functions are non-tractable (non-polynomial) only in respect of some small aspects of the input (called the input parameter) may suggest that a number of cognitive capacities that are considered to be computationally non-tractable may be otherwise tractable in respect of a portion of the input size. This may also support the idea that it's not only the quantity of the input that constraints problem tractability, but also the quality of the input instances and, then, the theoretical and empirical criteria that define the available input contained in the input size. In the end, such a theory also supports the idea that in order to tractably compute an output not taking into consideration the complete input size not only doesn't constitute a problem, but it's actually may be the reason for tractability itself.

Attribution of propositional contents may, in this sense, sufficiently on a portion of the input size (observable behaviour) and still be successful. Moreover, evidences such those that suggest the presence of a shared tendency to interpret the world in a dualist fashion may advocates for good candidates of mind-shaping processes for overcoming heterogeneity and also avoiding, performance-wisely, any completeness requirement. In the next section, we will take an in-depth look at how homogeneity and mindshaping influence the tractability mindreading and check if formal results actually corroborate further the hypothesis about the computational tractability of mindreading.

[78] Tha stands for Fixed Paramenter Cognition.

6.2 Does mindshaping make mindreading tractable?

In philosophy and psychology of social cognition is often stated that the cultural and social context has a deep influence on mindreading (MR). Among the available theories the Mindshaping (MSh) hypothesis claims that computational tractability of MR is reached through mechanisms that make the social domain homogeneous. We will break down this core claim of MSh and investigate the possible influence of homogeneity on MR tractability. To do this, we take action understanding as a case-study for MR. This enables us to bridge the gap between informal claims and formal (in)tractability results, by operationalizing Mindshaping homogeneity in different ways. We aim to achieve three results. First, we illustrate how to ground the claimed effects of mindshaping in formal computational complexity analysis. Second, we show how this can reveal new ways of interpreting homogeneity. Third, by bridging the gap between informal and formal theory, we propose a way to formally evaluate theoretical claims about the potential effects of cultural and social context on MR tractability.

6.2.1 Mindshaping, mindreading, culture ad tractability

Mindreading (MR) is considered, together with natural language, to be one of the definitory capacities of human cognition. It allows agents to explain and predict their own behaviour as well as that of others. However, MR is also influenced by the cultural context. It has been proposed that it may be culture specific (Adams et al. 2010), culturally inherited (Heyes and Frith 2014) and that this may also translate in better performances inside cultural groups (Perez-Zapata, Slaughter, and Henry 2016). Among these proposals

we find the Mindshaping hypothesis (Tadeusz W. Zawidzki 2008; Tadeusz Wieslaw Zawidzki 2013; Mameli 2001). This theory proposes that the actual contribution of the cultural context is to make MR possible in the first place. This is done by claiming that mindshaping (MSh) mechanisms complement MR so that the paradox of the intractability of MR can be solved. The paradox goes like this: while humans seem to be very good at reading other agents' intentions in a timely manner, theories of mindreading are computationally intractable (Alechina and Logan 2010; Apperly 2010; Tadeusz Wieslaw Zawidzki 2013) implying that cognitive capacities take unrealistic amounts of time to be computed (van Rooij, 2008). The intractability of MR is attributed to the problem of holism. This problem states that it is logically possible to associate any intentions to every behaviour, so that the intentions behind the actions of other agents will always be underspecified. However, also computational models of abduction corroborate MR intractability. Abduction is known for its computational intractability, meaning that the inferences postulated by these models require exponential amounts of time (Abdelbar and Hedetniemi 1998; Nordh and Zanuttini 2005; Bylander et al. 1991; Thagard 1993), unless several restrictions are introduced (van Rooij and Wareham 2007). The intractability of MR means that current theories cannot yet explain why people can do mindreading quickly. Rather than rejecting the theory outright, the paradox might be resolved by revising it instead. MSh claims to solve the issue by proposing that that behaviourally implemented cultural mechanisms "homogenize" the social environments and shape agents in order to be easily interpretable. Unfortunately, Zawidzky fails to provide any formal analysis in support of his informal claims, leaving a conceptual gap between philosophical and formal analysis unexplored.

In the remainder of the paper we will propose how operationalizing MSh core claims into possible ways in which special case of mindreading, viz. action understanding as modelled by Bayesian inverse planning (C. L. Baker,

Saxe, and Tenenbaum 2009a; C. L. Baker, Tenenbaum, and Saxe 2008) can be restricted. Our argument is threefold. First, we show how MSh core claims about the influence of mindshaping mechanisms can be reduced to homogeneity of MR. Second, after showing that action understanding can be considered as sub-capacity of MR consistent with MSh, we operationalize MSh core claims into parameters for action understanding (Blokpoel et al., 2013). Finally, we show that some Msh-based restrictions on action understanding lead to tractability of MR, but others do not. Furthermore, by operationalizing homogeneity, we discover new ways of understanding the consequences of MSh assuming that MSh leads to tractable MR. While we do not claim that our operationalization provides an exhaustive picture of the effect of culture on MR, we aim to propose a way to reach formal clarification and methodology for evaluating how computational tractability can help in individuating aspects in need of theory revision.

6.2.2 Reconstructing Mindshaping

First proposed by Mameli (2001) and later developed by Zawidzki (2009; 2008; 2013), MSh proposes that the success of human social cognition is explained by the fact that human mindreading is complemented, and not substituted, by a set of mindshaping mechanisms that "shape our socio-cultural environment in ways that make coordination exponentially more tractable" (Tadeusz Wieslaw Zawidzki 2013, p.9). Otherwise, for the reason that we have seen above, MR is computationally intractable. Example of mindshaping mechanisms include: goal imitation, cognitive imitation, overimitation, the chameleon effect (Chartrand and Bargh 1999), pedagogy, norm following and self-constituting narratives"(T. W. Zawidzki 2013, p.144). What all of abovementioned phenomena share, according to MSh, is the characteristic of implementing social expectancy and conformity mechanisms.

These phenomena are at the base of the behavioural processes that "mindshape" the social environment interpreted through MR.

MSh takes a precise stance on the evolution of human social cognition, here instead we will focus exclusively on the consequences that MSh has over the efficiency of full-fledged human mindreading. Mindshaping mechanisms are hypothesized to positively affect the reliability of mindreading (Tadeusz Wieslaw Zawidzki, 2013, pp. 146-147). Reliability, however, is used by Zawidzki in two different (orthogonal) senses: accuracy and tractability. Insofar as explaining how mindreading can be tractable, the accuracy-sense of reliability is not relevant. It is possible that an intractable function is extremely inaccurate and it is also possible to have a tractable function that is accurate.[79]

One of the principal feature of the MSh hypothesis is the separation between MSh mechanisms and MR. According to MSh, the tractability of MR is not caused by directly modifying the workings of this capacity. MSh mechanisms are not components or modules (Zawidzki 2009; 2013) of an higher-level mindreading capacity. They are instead complementary to mindreading and work by modifying, or shaping, the socio-cultural environment that can be considered the actual input and search space of MR. In MSh the solution to mindreading intractability "lie(s) not *within* human mind readers, but, rather, *outside* of them" (Zawidzki, 2009, p.5). That is also reflected by the inherently relational and social nature of mindshaping, an aspect that can be easily linked to the fact that "a mindshaping mechanism is one that aims to make a target's behavioral dispositions match, in relevant respects, some model" (Zawidzki, 2013, p.86). Where the model may be a real

[79] The issue of optimality is often used to associate intractability with accuracy, where it is claimed that intractability is caused by optimality. However, it is known that approximation (a relaxation of the optimality constraint) does not necessarily grant tractability. (van Rooij and Wareham 2012)

as well as virtual entity like a role-model or an abstract normative element. Through processes of socio-environmental pressure MSh's mechanisms of conformity are selected and reinforced. Furthermore, the key elements of such processes will lead us to the last step in our deconstruction, the actual effects of MSh mechanism on MR tractability.

Msh implements conformity mechanisms that make the agents in the social environment behave accordingly to a normative model of some sort. Zawidzki calls this conformity homogeneity, and considers it the primary cause of tractability of MR (REF). However, given his informal characterization of MSh, it is imperative to break down the various way in which "homogeneity" may be intended. The first elements to assess is that homogeneity takes two forms in MSh:

- Cognitive Homogeneity (CH): the form of homogeneity that comes from the common nature of the human cognitive system. This kind of homogeneity is considered responsible for the fact that humans share the majority of their propositional attitudes;
- Mindshaping Homogeneity (MH): the form of homogeneity that comes from MSh mechanisms. This form of homogeneity is implemented in the socio-cultural environment at has the role of mitigating the variations that are brought by elements such like experiential history, motivation, attention, and memory.

These two forms of homogeneity are not equally represented in the MSh framework, where the balance is shifted toward the second type. Here the claim is that MR tractability cannot be reached only with CH. MSh homogeneity is claimed to be a necessary condition for MR tractability. To

investigate this claim we have to further refine these two types of homogeneity and see how they can or cannot lead to tractability.

6.2.3 Operationalising homogeneity

As we have seen, Zawidski claims that MR is made tractable by a set of homogeneity implementing behavioural mechanisms. However, tractability is a formal propriety of functions and cannot be intuitively demonstrated. A way to clarify the influence of homogeneity may come from functionally defining MR in the following way:

MR: a mapping from an *observed social environment* (consisting of observed behaviours, actions and context) and *social knowledge* to *intentional attributions*.

It is known that some intractable functions can be tractable for a subset of their input domain (Downey & Fellows, 1999; van Rooij & Wareham, 2008). These subsets can be defined as restrictions on properties of the input. When the tractability of a function is obtained through such restrictions it is said to be fixed-parameter tractable for that subset of the input. Likewise, homogeneity's influence on MR can be interpreted as in the following ways: As a restriction on the number of different occurrences; or as a limitation on the variability in a set. This latter sense can be linked from the start to cognitive homogeneity, since it seems to be concerned more with sharing intentions than restricting their number. The following are the ways in which, we propose, homogeneity can be interpreted when put in the context of MSh influence on MR:

- Having a restriction on the number of available behaviours in a population;

- Having a restriction on the number of available intentions in a population;
- Having a less ambiguities in the links between behaviours and intentions due to social knowledge.

3.

However, the exact nature of these restrictions will remain unclear until MR is not formally individuated. Currently cognitive science has no model for generic MR available. There are, however, computational models of sub-capacities of MR. Here especially we will refer to human action understanding as modelled though Bayesian Inverse Planning (BIP). According to the BIP model, action understanding can be seen as a form of inverse planning, an abductive inference from actions to goals (C. L. Baker, Saxe, and Tenenbaum 2009a) guided by the *principle of rationality*[80]. The inferential nature of BIP is furthermore consistent with the version of MR that Zawidski's MSh addresses. In his account, in fact, he refuses modularity, accepting by consequence the isotropy and abductive nature of MR (J. A. Fodor 1983; Heal 1996). Although, Zawidski acknowledges fast & frugal heuristics and simulation theory accounts of MR, his concerns for intractability of MR are based on the problem of holism and thus pertain to abduction-based MR. Bayesian-inverse planning, although a sub-capacity of generic mindreading, makes for an acceptable first formal investigation of Zawidski's claims given its abductive nature.

Taking BIP as our case-study we can now investigate the possible input-restrictions that can result from MSh for action understanding, and show which of these restrictions obtain tractability of this MR sub-capacity. The BIP model can be informally characterized as follows:

[80] "the expectation that intentional agents will tend to choose actions that achieve their desires most efficiently, given their beliefs about the world" (Baker et al., 2009, p. 2)

BIP: A mapping from a set of observed actions A, observed states S and probabilistic relations between actions, states and goals G to the most probable goals G given A and S.

Blokpoel et al. (2013) have provided intractability results for several input-restrictions of the BIP model. Such input restrictions take the form of parameters for the following elements of the BIP model:

- The number of actions |A| that are observed by an interpreter;
- The number of goals |G| that are inferred by an interpreter;
- The number of available actions *a*;
- The number of available goals *g*;
- The inverse probability of the most probable goals *1-p, this is related to the probabilistic relations between actions, states and goals*.

While none of the above restrictions are by themselves sufficient for reaching tractability, there are combinations of them that gave the following intractability results.

1. {|A|,a,|G|} is fixed-parameter intractable;
2. {|A|,a,g} is fixed-parameter intractable.

So, restricting |A|, |G|, g and a, by themselves cannot lead too tractability of action understanding. Since no results are known for 1-p by itself, it is safer to assume that restricting 1-p by itself also does not lead to tractability of action understanding. However, if the right parameters are restricted, then action understanding is computationally tractable. The following two results show that if either (1) or (2) or both conditions hold, then action understanding is tractable.

1. {|G|,g} is fixed-parameter tractable;
2. {1-p,g} is fixed-parameter tractable.

The above-mentioned parameterizations of BIP, the intractability results and the previously given interpretations of homogeneity makes now possible to operationalize homogeneity as we have previously defined it. Here we go beyond what is currently in the literature fleshing out more detailed effects that mindshaping might have, compared to given an overview of the mechanisms.

- Restricting the number of observed actions |A|:

  |A| defines the number of candidate actions that an interpreter is actually following in order to infer the relative goal. The socio-cultural context can be seen as influent on the way action are performed, but also on the prescribed action for a given situation. This restriction can be seen as working in tandem with |G|, but even in combination with it tractability cannot be obtained.

- Restricting the number of inferred goals |G|:

  If action understanding is to be tractable, then one option is for |G|, the number of possible goals that an interpreter actually pursues, to be small (together with g). If, within a social community, an actor would like his/her actions to be timely interpretable to others, then this actor might pursue few goals at a given time so as to make |G| small. This behaviour might be the result of MSh mechanisms such as habitualization, culture and ritualization resulting from MSh.

- Restricting the 'complexity' of single actions *a*:

a can be seen as the maximum number of possible actions that are available at every point in time. This number is upper-bounded by the total number of possible actions that are available to a person. One way to interpret homogeneity is that *a* is limited in a certain population as a result of MSh.

- Restricting the 'complexity' of single goals *g*:

  g can be seen as the maximum number of possible goal attributions available for the given inference. This number is upper-bounded by the total number of intentions available to an agent. It is claimed that CH may result in a population where people's sets of intentions overlap a lot (i.e., they share most of their possible intentions). This, however, does not necessarily restrict the size of that set and consequentially, it does necessarily restrict *g*. If anything restricts g, it seems there must be some not yet clearly defined MSh or MR process that does so. Due to the ubiquity of *g* in tractability results discovering these processes would be paramount for having a complete picture of the relation between homogeneity and tractability of MR.

- Restricting the relations (i.e., probabilistic dependencies) between variables and the prior probability of variables such that the most probable goal attribution has a high probability, i.e., 1-*p* is low:

  The relational probabilities between variables can be seen as encoding the social knowledge that is brought to bear when inferring the most probable goal. The prior probabilities of variables can be seen as the disposition a person has towards particular unobserved variables (such as goals) at the time of the inference. This knowledge can, e.g., be shaped by pedagogy, norm following and imitation.

The considerations made above show that Mindshaping homogeneity can be operationalised in many different ways. By themselves, the individual

operationalisations do not lead to tractability of mindreading. As we have seen multiple operationalisations need to be in effect simultaneously in order to render mindreading tractable, viz. g *and* |G| or g *and* 1-p. These two positive results, when linked with homogeneity interpretations, also reveal a gap in the array of otherwise consistent MSh effects. While one of the main claims of MSh is the importance of MH for the tractability of MR, the here considered tractability results show that a restriction on *g* is always necessary for MR tractability. The fact that no obvious link between the two types of homogeneity have been found with *g* indicates then a that a challenge to explain how a restriction on *g* is reached still remains.

6.2.4 Discussion

It is clear by now that multiple restrictions on action understanding are consistent with at least a part of the MSh claims. Even for a restricted case of mindreading such as action understanding, some of these restrictions have an effect on the tractability of this capacity, while others do not. This has important consequences for claims regarding the effect of homogeneity on the tractability of mindreading. For those restrictions that are consistent with MSh claims (viz. as |A|, |G|, a and 1-p), we now have formal proof which of these claims do and do not lead to tractability of a sub-capacity of MR. The same methodology can be applied to different computational models of MR capacities to further substantiate these claims. For those restrictions that are not (yet) consistent (viz. g) with MSh claims, these may point to yet undiscovered effects (and mechanisms) of mindshaping worth further investigation. In this way our analisys does not only provide a formal assessment of the claims about the effect of mindshaping on the (in)tractability MR but also provides a methodology that may lead to the discovery of new mindshaping effects and mechanisms.

Chapter 7

The minimal complexity hypothesis

In philosophy of mind and computation an age old question still stands: what makes a physical object sufficiently complex (Searle 1990) to implement computation? Having an answer is even more important now that we have a mechanistic proposal (Piccinini 2009) and the necessity of explaining concrete computation is felt as more urgent. In order to reach a satisfactory answer we sure need an adequate notion of computation (Fresco 2011), but it can be argued that is also important to possess an explanatory adequate and rigorous notion of complexity.

Looking at how cognitive science came to terms with the bounded nature of the human cognitive systems will be of use in reconstructing this complexity notion. Here we take a look at proposals like "the bounded brain hypothesis" (Cherniak 1990) and "the tractable cognition thesis"(Van Rooij 2008). The first grounds the bounded nature of the human cognitive system in quantitative neuroanatomy. The TCT employs instead both a rigorous notion of complexity and the computational framework. This complexity notion, borrowed from computational complexity theory, is utilized for demarcating between possible and plausible cognition. However, the complexity of a computational problem is usually considered insensitive to differences in computational models and their implementations (Van Rooij 2008). This complexity notion, therefore, makes the TCT clash with the intuitions that grounds the BBH, that is: the capacities of a cognitive systems, much like those of an artificial computer, must be somehow constrained also by the performance of the concrete computational machinery.

The TCT acknowledges the boundaries that the neuronal machinery poses on computational cognition only implicitly, but it mainly focuses on searching for the upper boundaries of cognitive/neural computation. However, in order to catch what makes a "physical object sufficiently complex" we need a kind of "minimal complexity" that accounts for the set of necessary and sufficient properties that makes the computational transition possible. In this case, we will look for complexity but the other way around, starting then from the implementation level.

One example may come from Tononi's notion complexity (Tononi 2008): a measure of the architectural properties necessary for a cognitive system in order to express a certain capacity. This notion of complexity captures a useful notion of "critical mass" for cognitive systems but only focuses on functional and organizational properties. A complexity evaluation should, however, not be limited to evaluating architectural features but should also consider efficiency and performance as relevant. That becomes evident if we look at the evolution of manmade computers, where the growth in performances and then efficiency played a big role.

A notion of minimal complexity that accounts for the architectural articulation of systems and also for the efficiency that it reaches in expressing its capacities may avoid a pan-computational drift while not being too restrictive. Furthermore, it may help in bridging the gap between the upper two level of Marr's hierarchy and the implementation one.

7.1 Setting the ground

Most of the struggle, and the hunger for further developments, in the current debate on the Computational Theory of Mind (CTM) derives from the desire to be as much as possible inclusive in the definition of "computation". This tendency translated in two major trends inside the CTM. On the one

hand, we had the progressive broadening of the actual notion of computation adopted (Piccinini and Bahar 2013). As a matter of fact, contemporary computationalism has long abandoned the strong analogy between cognitive computation steps and state transitions of Turing Machine (Putnam 1975; J. Fodor 1975) and instead considers cognitive computations as "any process whose function is to manipulate medium-independent vehicles according to a rule defined over the vehicles" (Piccinini and Scarantino 2010). On the other hand, such an evolution calls for the definition of a criterion that makes possible to pinpoint the actual type of computations that cognitive systems have developed and the characteristics of the neuronal machinery that implement them. It is in order to respond to these two necessities that the mechanistic approach (Piccinini 2007; Fresco 2014; Piccinini 2015) introduced the distinction between a computational layer and a concrete functional layer. The latter of which takes the form of a system of functionally organized components (a mechanism) that possess all the relevant properties in order to implement computations. This has two main consequences for the CTM: it detaches the computational explanation from the functional ground whilst making possible to implement the same computation in an array of different concrete computing systems; it avoids an escalation toward pancomputationalism by considering as cognitively relevant only those systems that are mechanisms in the first place and that have the function of manipulating medium-independent vehicles[83] by the means of an appropriate rule-set .

However, if we set aside for a moment the considerations about what is possible in principle, we have to acknowledge that there are only two

[83] The requirement of medium-independency excludes an eventual mutual dependency between the intrinsic characteristics of the computational mechanisms and a certain computational model. This way the case of a single mechanism implementing only a single model of computation is ruled out.

concrete computational systems that really deserved our attention so far: man-made computers, and nervous systems. In reflection of that the sometime tacit, core of most debates, is whether brains and ordinary computers share similarities under the standpoint of computation or not, and in the first case to what extent. There have been certainly other entities for which, sometimes, a computational nature has been inquired like cells, proteins, of even piles of sand, but all these cases are derivation of speculative pancomputational drifts, rather than objects of genuine computational analysis. In this sense while it is possible to have a computational description for such entities they cannot be considered computational in the full sense. So, there is still space for asking what are the precise features that make this demarcation possible and if computational systems possess some intrinsic properties that distinguish them from non-computational mechanisms. However, it's worth considering that when considering the two cases of computational system indicated above, a striking element of similarity, in our opinion, has been overlooked: the common medium offered by electricity.

Following this intuition, we will claim that electrical elaboration should be considered as a fundamental ingredient of computational systems. In this perspective, the system, in order to make the leap from mechanical to computational system, has to possess the functional and structural features to support the much more efficient and flexible transmission of information through electricity. Also, we propose that the importance of electrical elaboration calls for an update of the criterion that separates computational and non-computational mechanisms. This update should be developed having in mind the following claim: in addition to the architectural requirements of a computational system we should also ac- count in the CTM for those crucial performance advantages that are consequence of the appearance of electrical elaboration.

In order to support this hypothesis, we will see how electrical elaborations appeared in both biological organism and man-made systems and what kind of advantages it has brought to the both of them. In the end we will propose that is possible to synthetically express the minimal set of feature necessary for a system in order for it to solve a computational problem by defining a form of ”minimal complexity” that accounts for the influence that certain performance characteristics have in relation to the capacities that said system can express.

7.2 From mechanosensory to electrical elaborations in organisms

It is not easy to mark a sharp transition from mechanical to electrical solutions in nature, since electricity is the key element that sets animals apart from all other living organisms, primary by offering them the inestimable advantage of motion. Even if electricity in animal kinesis is not the driven force, as in electrical motors, it is involved in triggering mechanical contractions.

There are profound differences between electrical (and electronic) artifacts and animals. Man-made electrical power is mainly conveyed through metallic conductor. Computers, the artifacts managing electricity at a level of sophistication often compared to the brain, are made of semiconductors, such as silicon and germanium. Nature opted for the only electrical conductors compatible with organic materials: ions.

The biophysical breakthrough of exploiting electric power in animals has been the ion channel, a sort of natural electrical device, whose details have been discovered only recently (Neher and Sakmann 1976). It allows the flow of a specific type of ion only, across a cellular membrane, and under certain exclusive circumstances only, typically being the difference in voltage between the internal and the external areas of the cell. The first ion channel to appear in evolution was the potassium K^+ channel, which appeared about

three billion years ago in bacteria. It evolved into the calcium Ca^{++}permeable channel in eukariotes, and finally into the sodium Na^{+} channel, already found 650 million years ago in both ctenophora and early bilateria (Zakon 2012). From then on sodium channels became a widespread solution.

The calcium ion appears as the electrical carrier shared between the two strategies of offering kinesis, by contraction, and giving control, with neurons. A theory has been proposed for the Ca^{++} ion channel, that unifies the origin of contraction triggering and neural signaling, as evolved responses to the ancient emergence reaction of cells to external calcium influx after membrane damage (Brunet and Arendt 2016). Intracellular calcium is highly toxic because it forms insoluble precipitates with phosphate, therefore all eukariotic cells are well prepared to detect calcium ions. The two major repair strategies are the contraction of an actomyosin ring around the local channel of calcium influx, and the exocytosis of vesicles that seal the damaged membrane. The first mechanisms evolved in muscle contraction, the second in neurochemical transmission at synapse.

The key role of these two ions is confirmed by comparisons between extant species without specific Na^{+} Ca^{++} ion channels, such as fungi, and animals with simple nervous systems, such as ctenophora and sea anemone (Liebeskind, Hillis, and Zakon 2011), but many, if not most, of the details are still uncertain, and inextricably linked to a better understanding of the phylogeny of metazoans. Moroz (2009)proposed a hypothesis for the phylogeny of the neuron independent from the history of ion channels. One shared prerequisite of all neurons, from a genomic standpoint, is the capacity to express many more genes and gene products than other cell types. In fact, other cells can also exhibit massive gene expression, as a result of severe stress responses, and typically before death. Neurons might have evolved in ancestral metazoans from other types of cells, as the result of development in

the adaptive response to localized injury and stress, which gradually stabilized in cells supporting and maintaining the expression of multiple genes and gene products in normal conditions.

Sodium and calcium ions dominate the scene in the brain. There is a typical di↵erential distribution of ions inside and outside the neuron, with higher concentrations of K^+ inside, and Na^+, Ca^{++} and Cl concentrated more in the extracellular space. A characteristic of almost all neurons is an internal negative potential at rest, typically of -40 to -90 mV, due to an excess of negative charges with respect to the outside of the cell. This is due to the presence of organic ions, too large to leak across the membrane, and because potassium-permeable channels allow a continual resting efflux of K^+.

The minimal equipment for electrical control of animal kinesis is by motoneurons and sensory neurons. Motoneurons send signals to muscles that are transformed into mechanical actions. Acetylcholine, the most common neurotransmitter (Dale 1935), is released upon action potential by the motoneuron, and bind to nicotine acetylcholine receptors in the junctional folds of the muscle fibres, producing and end plate potential of the muscle junction. This potential, in turn, activate voltage gated Na^+ channels in the muscle fibres, causing influx of Ca^{++} ions in the T-tubules, sort of folding in the muscle membrane. This way the circle is closed, with the contraction reaction to Ca^{++} influx, described above. For the animal to move on purpose, motoneurons should be activated by a minimal sensation of the environment, which is provided by sensory neurons, receptors that transduce a specific form of energy (chemical, mechanical, thermal, acoustic, light) into a change in membrane potential (Zigmond and Bloom 1999).

Instead, neurons like interneurons are the basis of the electrical computation in the organisms, and both their dendrites and axon connect with other neurons, rather than muscle or receptors. The synapse is the

additional critical element to make a computational use of electricity. It is the synapse that allows neural system to be plastic, which is the secret for their ability to perform a huge range of computational functions (Bermúdez-Rattoni 2007; Blumberg, Freeman, and Robinson 2010). The eminent role of the synapse in intelligent behaviour has roused interest on its origin and evolution, and motivated explorations even more challenging than those on ion channels and the neuron. A kind of ancestor of all synapses, the ursynapse, appeared in choanoflagellates about one billion years ago, evolving in protosynapse similar to the extant ones in cnidarian, around 700 million years ago (Ryan and Grant 2009).

With the growing of the computations performed between sensory neurons and motoneurons during natural evolution, it became more convenient to assemble together the computational part, with respect to the electrical part strictly related to mechanical control. This evolution can be traced in metazoans, with the ancient diffuse nerve nets of cnidarians and ctenophores, the bilobed ganglia in polyclade flatworms, more complex multiple cephalic ganglia in many gastropod molluscs, and complete brains in vertebrates (Roth and Dicke 2013).

Mechanics and electricity are closely interwoven in all animals, still it is possible to conceptually distinguish the contribution of an electrical computation, in place of a mechanical approach, to a range of ordinary problems faced by the animal. Let take the very common and general problem of exploring the environment for rewarding places, for example food locations, for shelter, or mating. From the current position X of the animal there can be a set of new positions $Y_1, Y_2, \ldots, Y_n$ potentially useful. The pure mechanical solution is to move from X in turn to each of the Y_i unknown positions. Once reached, each position can be tested by the most primitive contact sensors, mechanical or chemical. Energy can be saved if the electrical

computation can screen out of the set $\{Y_i\}$ a smaller subset of candidate places for which it is worth paying a visit.

A first way to accomplish this computation is by advanced high-resolution vi- sion, which extends greatly the horizon of exploration beyond the contact space of the organism. Vision evolved from early phototaxis, the directional movement along a light vector towards (positive) or away from (negative) a light source. Marine larvae of animals with a pelagic-benthic life cycle use positive phototaxis to migrate upward in the water column, or negative phototaxis to migrate towards the benthic zone (Randel and Jekely 2016). Advanced scanning vision, common in vertebrates, cephalopods, arthropods and alciopid polychaetes, involves a tremendous amount of computation, for depth perception, shape analysis, invariant recognition (Palmer 1999). The advantage stands in the possibility of selecting, in the exploration problem, a very narrow set of regions that are worth to explore mechanically.

A further step in substituting electrical computation to mechanical strategies in environment exploration, is by storing previously visited places, and using appropriate algorithms in matching observed cues with memories of places. This is the innovation given by hippocampal place cells (O'keefe and Nadel 1978). An even more subtle cognitive computation can be applied by animals able to remember the episodes linked with place, in screening which one is worth to visit. Scrub jays make use of a form of episodic memory, that allows them to remember not just the places where they stored food, but also the time elapsed between caching and recovery (Clayton and Dickinson 1998). If the time is enough for food to degrade and become unpalatable, scrub jays use this computed information to save mechanical energy, avoiding the visit to caches of degraded food.

A different problem can be the selection of the tool appropriate for performing a certain action. In presence of a set $\{T_1, T_2, \cdots, T_n\}$ of tools, only one of which is appropriate for action X, the mechanical solution is to try in sequence all tools until the correct one is found. A better way is to run electricity in brain neurons, in mental simulations of the combination of a tool T_i with the action X, and decide which is correct, or at least shortlist the best candidates. Jane Goodall reported several cases of chimpanzees choosing directly sticks appropriate (in length and shape) for termite fishing, without overt behavioural trial and error (Goodall 2010). Julia, a chimpanzee raised in humanlike cultural environment, was able to determine mentally what kind of key was needed to open a locked box, without overt trial and error (Döhl 1968).

7.3 From mechanical to electrical elaborations in man-made systems

Contrary to what happened for organisms following the transition from mechanical to electrical based human computing system, being a recent and well known history, is much more straightforward. Almost all the history of mathematics is scattered with inventions of mechanical devices helpful for some sort of operation (Goldstine 1980). Due to the influence of philosophical mechanicism, from 17th to 19th century a plenty of automata was built. For instance, Jacques de Vaucanson was popular in France for his mechanical creations, like the Tambourine Player, the Flute Player, and the Digesting Duck. The most acclaimed in history inherited their fame from the notoriety of their inventors, like the Pascaline, developed in 1645 by Blaise Pascal (Kistermann 1998). This machine was able to compute additions and subtractions of two numbers up to 8 digits. It did so by counting the number of rotations wheel, each corresponding to a digit, with a lever mechanism that takes care of a carry. Some thirty years later Gottfried Leibniz had the

conceptual idea for a machine that extended the calculations of the Pascaline, performing the operation of multiplication by repeated additions using a sliding carriage as counter and stepped-drum to store the multiplicand setting.

The breakthrough from mechanical calculation of the basic arithmetic operations to general mechanical computations is due to Charles Babbage (Babbage and Babbage 2010). His first development was the Difference Engine, aimed at producing numerical tables of arbitrary polynomials up to degree six. The strategy was to use the values of the derivatives of various degrees at steps of integers of the polynomial variable, as in the Newton's method of divided differences. This way, only the first value of the polynomial had to be calculated, the table of all other values can be constructed using additions only. Sadly, Babbage during his lifetime failed to actually construct any of the many machines he conceived and designed. Certainly, one reason was the extreme challenge of the mechanical construction beyond the technologies of that time. His second version of the Difference Engine was completed in 1991 at the Science Museum of London (D. D. Swade 2005). The machine consisted of 8000 mechanical parts and weights 5 tons. Its core is a set of 14 twin cams arranged in a vertical stack. Each of 14 cams has a companion cam the profile of which is a geometric inversion of its mate. The 28 paired cams control the lifting, turning, and sliding motions required to execute the repeated additions for the Newton's method. This machine calculates and tabulates any seventh-order polynomial to 31 decimal places.

Even if polynomials up to degree six are a powerful approximation of most useful mathematical functions, the Difference Engine is not yet a generic computer, a transition Babbage made in 1833 with the project of the Analytical Engine (Goldstine 1980). The design of this machine included programmability, by punched cards, the separation between the mill, the core

processing unit performing basic operations, and the store, in which the variables to be operated upon, and the results, are placed. All these functions, realized mechanically, were vastly more demanding than those for the earlier machine. The mechanisms for direct multiplication and division, required complexities well beyond those for the repeated additions in the Difference Engine. In 2010 Graham-Cumming launched Plan 28, a campaign to raise funds to build the Analytical Engine, no actual construction had been yet started.

Without a physical exemplar of such machine it is di cult to discuss the limits of a mechanical man-made computation. From the point of view of the performances, Baggage achieved a capacity of 1000 numbers of 50 decimal digits for the store, and a speed of the mill of one multiplication per minute, and one addition per second. It is also problematic to assess the class of computational function the Analytical Engine can perform. In theory, its mill can form the basis for multistate logic, with logic state being physically manifested as a spatial configuration of the functional parts themselves, and state changes orchestrated by parts displacement.

Reif and Sun (Reif and Sun 2003) conceived a mechanical system based on frictional contact linkages between components, instead of toothed gears, and sketched a mechanical system of rigid objects whose surfaces are composed of patches specified by rational coefficients. All objects interact by surface contact, and each patch can behave as either purely frictional or purely sliding. The system configuration is in term of all relative positions of the patches on the surface. The system evolves from an initial configuration to a final configuration. Reif and Sun went on to show that such a system can encode the configuration of the Universal Turing Machine. Of course, this is just an abstract demonstration, which can hold only if there is no error at all in the frictional and sliding motions. Still, it is interesting for our purposes, in that ideally a mechanical system may equate an electrical digital computer.

The important difference, as we will discuss later, is in the complexity of the problem that can be treated mechanically.

Short after Baggage, electricity was beginning to being put to practical use, with the first large-scale electrical supply networks in the States, followed by radio transmission at the beginning of the 20th century. But mechanical computing devices were not immediately replaced by electrical devices, and the mechanical period for devices such as hand-held calculators, had a long overlap with the electrical era (D. Swade 2011). However, soon electricity became the principle of attention when the aim was to move toward general computation, as already proposed by Babbage. In several intermediate solutions electricity was combined with moving parts, essentially in two forms.

A first form is interesting in resembling the primary use we found in animals of electricity for motion, rather than just computation. It is the case when the logic of the system is still purely mechanical, but driven by electrical motors. Howard Aiken's Harvard MkI is an example of such solution. Electromechanical devices are those in which electricity drive directly the logic, even if through a moving part, like in contact-relay switches. Konrad Zuse in German adopted this transition, while Z1, his first computer completed in 1938 was purely mechanical machine, the next one, Z2, combined a mechanical memory with an arithmetic unit made of 200 electromagnetic relay switches (Zuse 1982). Starting with ENIAC (Goldstine and Goldstine 1946), the computer built in 1943 at the Army Ordnance Department to quickly calculate ballistic missile trajectories in wartime, electricity began to sweep the board. Vacuum-tube, with no moving parts has a speed 1000 times that of switching relays. There is no way to compare in time the divergence between mechanical and electrical performances, a rough figure can be given for electromechanical devices, which are still in use as switches. A micro electromechanical relay switch nowadays is 1 million times

slower than a CPU gate, and consumes somewhat like 100,000 more power energy.

7.4 Minimal complexity

Recently the notions of complexity employed in cognitive science have followed two main approaches. One of these is Tononi's proposal of employing a measure of the complexity as measure of the properties necessary for a cognitive system in order to express a certain capacity. As we have seen in chapter five such a measure is based on Shannon's notion of information it increases as the system is capable of integrating enough information so that "the information generated by the system as a whole is more than the information generated by its part taken independently" (Tononi 2010). This notion of complexity, while capturing the key bug vague notion of critical mass for cognitive systems, does not have any relation with what the system does, what problem is it able to solve and, especially, why is it able to do that by a performance standpoint. It provides a useful insight on what kind of tool we want for measuring the overall architectural complexity. However, it ultimately fails to provide any insight about what problem such a system is able to solve and it focuses only on the capacity of expressing consciousness.

On the other hand, approaches like that of computational complexity offer a completely different point of view on what we mean by the term complexity. In chapter two we have seen how computational complexity is a concerned with the hardness of computational problems. It analyses how the computational resources (time and memory) rise in function of input size and classifies such problems in term of their tractability. Applying this theory to cognition produced "the tractable cognition thesis". Through this thesis a new constraint is imposed on computational cognitive systems: cognitive

computations don't have only to be effective (computable) but also efficient and therefore tractable. The hardness of the problem becomes in a way an indirect measure of the effort taken by the (cognitive) system to solve such problem. If a certain model of a specific cognitive capacity presupposes an intractable computational problem, then said model should be also considered explanatorily unfeasible.

A couple of remarks should be made at this point. Introducing the tractability constraint inside the cognitive frameworks partially answers to our previous criticism against Tononi's complexity. It provides an indirect way to account for the cost that the cognitive system has to pay in order to perform a certain function or implement a certain capacity. Cognition is by consequence no more considered as an unbounded phenomenon but has to come on terms with the limits of what can be conceived as computationally feasible. It should be anyway noted that while tractable cognition recognizes the bounded nature or cognition, it also acknowledges the invariance thesis and thus ignores the influence that the neuronal machinery may have on the capacity of solving a certain computational problem.

What if, instead, we would like to explore the intersection between the set of lower level-mechanisms (Milkowski 2013) and that of computational mechanisms by approaching it from the lower bound? In that case, we will still be looking for a complexity notion, but of a different kind from the ones that we have already considered and analysed. This notion of complexity would be one that tries to catch the very characteristics that makes the computational transition possible in the first place. In fact, if the two previous approaches to complexity share something is that they both try to pinpoint a way to define what does possessing a certain set of relevant features means for a system. However, what the appearance of electrical elaboration in organisms, the improvements that it brings in them and its ubiquity in the evolution of man-made computers seems to suggests is that accounting only

for the complexity of the architecture may not be enough when we look for this minimal set. We have to retain the insight on the bounded nature of cognition that tractable cognition provides adding to architectural requirements such as Wimsatt's aggregativity (Wimsatt 1997) certain conditions on what kind of processing is actually executed and how such processing takes place. What the history of electrical elaboration points out is that the expression of a certain complexity level is not only limited to architectural feature but is actually pretty sensitive to the performance of the concrete computer too. That is indeed controversial because computation is often considered as medium-independent but that should not really come as a surprise. Architectural and performance-wise complexity are in reality deeply connected and that should be clear if we look at the second e third paragraph of the present chapter. Electrical elaboration presupposes the presence of the structural features needed in order to support electrical elaboration and the efficiency gains are a direct consequence of that, but not an epiphenomenon in respect to the capacity of the system. We're not proposing to abandon medium-independency but we should at least allow for a weakening of it in order to make possible to include a bit of performance sensitivity in the CTM.

However, in order to better define what we intend for minimal complexity we may find useful to borrow some elements from computational complexity. First and foremost, the focus on problems and what resources are needed in order to solve them. That is necessary because, contrary to man-made computers, there are no ways for estimating the computational power of organic systems. Minimal complexity much like computational complexity tout-court should then be measured indirectly. As we have already said computational complexity is mainly interested in lower bounds and takes a great deal of efforts in order to detach itself from the actual concrete computation. However, if we focus instead on those problems belonging to the complexity classes like L and P usually considered tractable, we may see

that the contribution of the system's performance may become more and more relevant as the problem be- comes simpler. That is true especially if the system is in evolution and therefore susceptible to leaps and jumps in its computational power due to the progress of its underlying machinery. Phenomena of this kind can be observed also in man-made computers. Since the beginning, computational complexity had part of its motivation in the awareness of the unique progress in performances offered by electrical computation. Hartmanis and Stearns published the first work on algorithmic complexity (Hartmanis and Stearns 1965) in 1965, the same year when Moore formulated his famous law on the exponential increase in hardware performances (Moore 1965). The deeply investigated division between complexity classes P and NP is also important because it allows to tell whether a problem, currently too di cult, can be solved in the future thanks to the expected growth in computer performances, or not (Papadimitriou 2014). So, in conclusion, we consider minimal complexity as a property of problems and, particularly as the class of complexity of the simplest problem(s) that are tractably solvable by a computational system, but not by a mechanism.

In the end, we can state that electrical elaboration suggests that mechanisms have to express a certain level of performance in order to be considered computational. The importance of that is captured by the minimal complexity that accounts for not only the overall architectural articulation of the systems but also the capacity of a system of solving certain problems for which certain performance characteristics are especially relevant. Including such elements becomes even more important if we consider that CTM is ultimately a theory about the nature of cognitive systems. The notion of minimal complexity is in fact perfectly neutral and when applied to the explanation of cognition, suggests that mechanisms have not only to express a certain level minimal complexity in order to be considered computational, but

also a certain level of minimal complexity in order to be considered computational and cognitive.

7.5 And then what?

Over the last decades, computationalism has been under attack by many critics [6] and has evolved in this regard. Connectionism, dynamicism, embodied embedded cognition and evolutionary robotics have been proposed as theoretical alternatives to classical computationalism. In CTM there are then two requirements to bring together: the seminal intuition, according to which cognition involves computations, and the risk for pancomputationalism, i.e., the idea that everything that's understandable in computational terms is a computer. While one can be skeptical about computations as they are described in classical computationalism, it must also avoid to throw the baby out with the bathwater. The above considerations aimed to explore the possibility of a fundamental computationalim. A kind of computationalim more grounded in the reality of neurocognitive computation, based on two basic constraints: the role of electricity in the evolution of computational cognition and the "minimal complexity" of that the system has to express in order to be able to perform its function. Taking into consideration these constraints, we are finally able to at least better frame the possible distinction those systems that perform computations (natural or man-made) and other kinds of mechanisms that may be only computationally describable. Furthermore, the new notion of minimal complexity, even if still at a preliminary stage, also suggest a novel, and pretty unorthodox way, of utilizing computational complexity for cognitive science.

Conclusions

We have started this thesis with the task of checking how complexity, and especially computational complexity, could improve our understanding of philosophically rich notions. We first looked at the philosophical and cognitivist relevance of complexity and of computational complexity after that. There we concluded that complexity is a notion in search of clarity and then we looked at how computational complexity could improve such situation by providing a rigorous but flexible way of characterizing plausibility. After that we moved on and shifter our focus on the applications that computational complexity may find in philosophy.

In the second section, we looked at philosophical applications. The first application that we considered is that with simplicity. There we've seen how computational complexity may be used to characterize one of two aspects of simplicity, parsimony, that can be used by a cognitive system as a principle of choice between contingent solutions. The behaviour of complexity into theories has been then taken into consideration in chapter four. There we have concluded that different theories can be evaluated through the notion of complexity that they imply. The next point of our analysis has been the comparison between computational complexity and Tononi's version of complexity. Through this comparison we've been able to see that different notions of complexity imply different theoretical approach and that the full cognitivistic explanatory spectrum needs more than one notion of complexity to be complete. Also, computational complexity has revealed to be more flexible than its counterpart, even if still more suite to capture notion like that of difficulty of execution and effort.

In the third section, we addressed more cognitively aimed applications. First we applied the tractable cognition thesis to mindreading in order to see how philosophical claims about this cognitive capacity can be re-evaluated and if they actually provide hints to ground formal analysis. After that we went in search of a minimal notion of complexity and, in order to do so, we proposed an unorthodox way of applying computational complexity to cognitive science. Instead of working our way by the upper bound, like tractable cognition does, we went the other way around and looked at how computational complexity can also provide hints about the implementation requirement of cognitive system.

Throughout our work we have achieved at least three goals: we assessed the philosophical relevance of computational complexity in cognitive science, we have seen how computational complexity scales and react when confronted with philosophical and cognitive notions. These are all relevant by a philosophical standpoint because they improve on our understanding of notions that can be applied to have an a priori evaluation of their validity. Furthermore, these results are relevant by a cognitivistic standpoint, since they provide a way of theoretically grounding technical solutions that, otherwise, would remain isolated in scope.

Further work may comprise applications of tractable cognition to cognitive capacities different than mindreading, like consciousness for example, where would be useful to look at the computational complexity of problems that are thought to be at the basis. The same can be done for more exquisitely theoretical notions like relevancy, or for addressing known problems like the frame problem.

Bibliography

Aaronson, Scott. 2011. 'Why Philosophers Should Care About Computational Complexity'. *In Computability: Gödel, Turing, Church, and Beyond...*, 58. http://arxiv.org/abs/1108.1791.

Abdelbar, Ashraf M, and Sandra M Hedetniemi. 1998. 'Approximating MAPs for Belief Networks Is NP-Hard and Other Theorems'. *Artificial Intelligence* 102 (1). Elsevier: 21–38.

Adams, Reginald B, Nicholas O Rule, Robert G Franklin, Elsie Wang, Michael T Stevenson, Sakiko Yoshikawa, Mitsue Nomura, Wataru Sato, Kestutis Kveraga, and Nalini Ambady. 2010. 'Cross-Cultural Reading the Mind in the Eyes: An fMRI Investigation'. *Journal of Cognitive Neuroscience* 22 (1): 97–108. doi:10.1162/jocn.2009.21187.

Alechina, Natasha, and Brian Logan. 2010. 'Belief Ascription under Bounded Resources'. *Synthese* 173 (2): 179–97. doi:10.1007/s11229-009-9706-6.

Alligood, Kathleen T., Tim D. Sauer, and James A. Yorke. 1997. *Chaos: An Introduction to Dynamical Systems*. Textbooks in Mathematical Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-59281-2.

Anderson, John Robert. 1990. *The Adaptive Character of Thought*. Psychology Press.

Apperly, Ian. 2010. *Mindreaders: The Cognitive Basis Of' theory of Mind'*. Psychology Press.

Arora, Sanjeev, and Boaz Barak. 2009. *Computational Complexity*. Cambridge, UK: Cambridge University Press.

Baars, Bernard J. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.

———. 2005. ‘Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience.’ *Progress in Brain Research* 150 (January): 45–53. doi:10.1016/S0079-6123(05)50004-9.

Babbage, Charles, and Henry Prevost Babbage. 2010. *Babbage’s Calculating Engines: Being a Collection of Papers Relating to Them, Their History and Construction*. Cambridge University Press.

Baker, Alan. 2013. ‘Simplicity’. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2013/entries/simplicity/.

Baker, Chris L., Rebecca Saxe, and Joshua B. Tenenbaum. 2009a. ‘Action Understanding as Inverse Planning’. *Cognition* 113 (3). Elsevier B.V.: 329–49. doi:10.1016/j.cognition.2009.07.005.

Baker, Chris L, Rebecca R Saxe, and Joshua B Tenenbaum. 2009b. ‘Bayesian Theory of Mind : Modeling Joint Belief-Desire Attribution’. *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society* 1 (2006): 2469–74.

Baker, Chris L, Joshua B Tenenbaum, and Rebecca R Saxe. 1995. ‘Goal Inference as Inverse Planning’.

———. 2008. ‘Bayesian Models of Human Action Understanding’. *Consciousness and Cognition* 17 (1): 136–44. doi:10.1016/j.cognition.2009.07.005.

Baker, Cl, and Joshua B Tenenbaum. 2006. ‘Modeling Human Plan Recognition Using Bayesian Theory of Mind’. *Mit.Edu* 1: 1–24. http://web.mit.edu/clbaker/www/papers/baker_chapter2014.pdf.

Beck, Jacob. 2012. ‘Do Animals Engage in Conceptual Thought?’ *Philosophy Compass* 7 (3). Wiley Online Library: 218–29.

———. 2013. ‘Why We Can’t Say What Animals Think’. *Philosophical Psychology* 26 (4). Taylor & Francis: 520–46.

Bermúdez-Rattoni, Federico. 2007. *Neural Plasticity and Memory: From Genes to Brain Imaging*. CRC Press.

Bertalanffy, Ludwig Von. 1968. *General System Theory*. New York: George Brazilier.

Blumberg, Mark Samuel, John Henry Freeman, and Scott R Robinson. 2010. *Oxford Handbook of Developmental Behavioral Neuroscience*. Oxford University Press New York.

Brunet, Thibaut, and Detlev Arendt. 2016. 'From Damage Response to Action Potentials: Early Evolution of Neural and Contractile Modules in Stem Eukaryotes'. *Phil. Trans. R. Soc. B* 371 (1685). The Royal Society: 20150043.

Bylander, Tom, Dean Allemang, Michael C Tanner, and John R Josephson. 1991. 'The Computational Complexity of Abduction'. *Artificial Intelligence* 49 (1–3): 25–60. doi:10.1016/0004-3702(91)90005-5.

Byrne, David. 1998. *Complexity and Postmodernism*. *Journal of Artificial Societies and Social Simulation*. Vol. 2. London and New York: Routledge. doi:9786610333837.

Carruthers, Peter. 2006. *The Architecture of the Mind*. Oxford University Press.

Carruthers, Peter, and Peter K Smith. 1996. *Theories of Theories of Mind*. Cambridge Univ Press.

Chaitin, Gregory J. 1966. 'On the Length of Programs for Computing Finite Binary Sequences'. *Journal of the ACM (JACM)* 13 (4). ACM: 547–69.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford, UK: Oxford University Press.

Changizi, M A. 2007. 'Scaling the Brain and Its Connections'. *Evolution of Nervous Systems* 3. Amsterdam, The Netherlands: Elsevier: 167–80.

Changizi, Mark A. 2001. 'Principles Underlying Mammalian Neocortical Scaling'. *Biological Cybernetics* 84 (3). Springer: 207–15.

Chartrand, Tanya L, and John A Bargh. 1999. 'The Chameleon Effect: The Perception–behavior Link and Social Interaction.' *Journal of Personality and Social Psychology* 76 (6). American Psychological Association: 893.

Chater, Nick. 1997. 'Simplicity and the Mind'. *The Psychologist* 10 (11). British

Psychological Society: 495–98.

———. 1999. ‘The Search for Simplicity: A Fundamental Cognitive Principle?’ *The Quarterly Journal of Experimental Psychology: Section A* 52 (2). Taylor & Francis: 273–302.

Chater, Nick, Mike Oaksford, Ramin Nakisa, and Martin Redington. 2003. ‘Fast, Frugal, and Rational: How Rational Norms Explain Behavior’. *Organizational Behavior and Human Decision Processes* 90 (1): 63–86. doi:10.1016/S0749-5978(02)00508-3.

Chater, Nick, Joshua B Tenenbaum, and Alan Yuille. 2006. ‘Probabilistic Models of Cognition: Conceptual Foundations’. *Trends in Cognitive Sciences* 10 (7). Elsevier: 287–91.

Chater, Nick, and Paul Vitányi. 2003. ‘Simplicity: A Unifying Principle in Cognitive Science?’ *Trends in Cognitive Sciences* 7 (1). Elsevier: 19–22.

Cherniak, Christopher. 1990. ‘The Bounded Brain: Toward Quantitative Neuroanatomy’. *Journal of Cognitive Neuroscience* 2 (1). MIT Press: 58–68.

Chomicki, Jan, and Jerzy Marcinkowski. 2005. ‘On the Computational Complexity of Minimal-Change Integrity Maintenance in Relational Databases’. In *Inconsistency Tolerance*, 119–50. Springer.

Clark, Andy, and David Chalmers. 1998. ‘The Extended Mind’. *Analysis*. JSTOR, 7–19.

Clayton, Nicola S, and Anthony Dickinson. 1998. ‘Episodic-like Memory during Cache Recovery by Scrub Jays’. *Nature* 395 (6699). Nature Publishing Group: 272–74.

Cook, Stephen. 2003. ‘The Importance of the P versus NP Question’. *Journal of the ACM (JACM)* 50 (1). ACM: 27–29.

Copeland, B Jack. 2002. ‘Accelerating Turing Machines’. *Minds and Machines* 12 (2). KLUWER ACADEMIC PUBL: 281–301. doi:10.1023/A:1015607401307.

Cordeschi, Roberto, and Marcello Frixione. 2007. ‘Computationalism Under

Attack'. In *M. Marraffa, M. De Caro and F. Ferretti (Eds.), Cartographies of the Mind, 37-49. © 2007 Springer.*, 37–49.

Coricelli, Giorgio. 2005. 'Two-Levels of Mental States Attribution: From Automaticity to Voluntariness'. *Neuropsychologia* 43 (2). Elsevier: 294–300.

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. Edited by Thomas M. Cover and Joy A. Thomas. New Jersey: John Wiley & Sons.

Crick, Francis. 1990. *What Mad Pursuit: A Personal View of Scientific Discovery*. Basic Books.

Cummins, Robert. 2000. 'How Does It Work?" versus" what Are the Laws?": Two Conceptions of Psychological Explanation'. *Explanation and Cognition*. MIT Press Cambridge, MA, 117–44.

Dale, Henry. 1935. 'Pharmacology and Nerve-Endings'. *Journal of the Royal Society of Medicine* 28 (3). SAGE Publications: 319–32.

Davidson, Donald. 1975. 'Thought and Talk'. *Mind and Language* 1975. Oxford: Oxford University Press: 7–23.

Davis, Martin. 2004. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Courier Dover Publications.

De Caro, Mario, and Maurizio Ferraris. 2012. 'Bentornata Realtà. Il Nuovo Realismo in Discussione'. *Torino: Einaudi*.

Dehaene, Stanislas, and Jean-Pierre Changeux. 2011. 'Experimental and Theoretical Approaches to Conscious Processing.' *Neuron* 70 (2). Elsevier Inc.: 200–227. doi:10.1016/j.neuron.2011.03.018.

Dennett, Daniel. 1978. 'Why You Can't Make a Machine That Feels Pain'. *Machinery* 38: 415–56. doi:10.1007/BF00486638Drescher.

Dennett, Daniel C. 1991. *Consciousness Explained*. Little, Brown and Co.

———. 1995. *Darwin's Dangerous Idea*. Simon & Schuster.

Dennett, Daniel Clement. 1989. *The Intentional Stance*. MIT press.

Dixon, James A., John G. Holden, Daniel Mirman, and Damian G. Stephen.

2012. ‘Multifractal Dynamics in the Emergence of Cognitive Structure’. *Topics in Cognitive Science* 4 (1): 51–62. doi:10.1111/j.1756-8765.2011.01162.x.

Döhl, Juurgen. 1968. ‘Über Die Fähigkeit Einer Schimpansin, Umwege Mit Selbständigen Zwischenzielen Zu Überblicken’. *Zeitschrift Für Tierpsychologie* 25 (1). Wiley Online Library: 89–103.

Downey, R G, and M R Fellows. 1999. ‘Parameterized Complexity’. *Proc 6th Annu Conf on Comput Learning Theory* 5 (1): 51–57. doi:10.1016/S1571-0661(04)00301-9.

Downey, Rodney G, Michael R Fellows, and Ulrike Stege. 1999. ‘Parameterized Complexity: A Framework for Systematically Confronting Computational Intractability’. In *Contemporary Trends in Discrete Mathematics: From DIMACS and DIMATIA to the Future*, 49:49–99. AMS-DIMACS Proceedings Series.

Dreyfus, Hubert L. 1972. ‘What Computers Can’t Do’.

———. 1992. *What Computers Still Can’t Do: A Critique of Artificial Reason*. MIT press.

Dunn, Michael. 2008. ‘Information in Computer Science’. In *Floridi L., Philosophy of Information*. Blackwell Publishing Ltd.

Edmonds, Bruce. 1995. ‘What Is Complexity?-The Philosophy of Complexity per Se with Application to Some Examples in Evolution’. *The Evolution of Complexity*. Kluwer, Dordrecht.

Enderton, Herbert B. 1977. ‘Elements of Recursion Theory’. *Studies in Logic and the Foundations of Mathematics* 90. Elsevier: 527–66.

Ferraris, Maurizio. 2012. *Manifesto Del Nuovo Realismo*. GLF editori Laterza.

Fodor, Jerry. 1975. *The Language of Thought*. New York: Thomas Crowell.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Edited by Zenon W Pylyshyn, William Demopoulos, Zenon W Ed Pylyshyn, and William Ed Demopoulos. *Philosophical Review*. Vol. 94. Theoretical Issues in Cognitive

Science. MIT Press. doi:10.2307/2184717.

Fortnow, Lance. 2004. 'Kolmogorov Complexity and Computational Complexity'. *Complexity of Computations and Proofs. Quaderni Di Matematica* 13.

Fresco, Nir. 2011. *Concrete Digital Computation: What Does It Take for a Physical System to Compute? Journal of Logic, Language and Information*. Vol. 20. doi:10.1007/s10849-011-9147-8.

———. 2014. *Physical Computation and Cognitive Science*. Vol. 12. doi:10.1007/978-3-642-41375-9.

Frixione, Marcello. 2001. 'Tractable Competence'. *Minds and Machines* 11 (3). Springer: 379–97. doi:10.1023/A:1017503201702.

Funtowicz, Silvio, and Jerome R Ravetz. 1994. 'Emergent Complex Systems'. *Futures* 26 (6). Elsevier: 568–82.

Gallese, Vittorio, Luciano Fadiga, Leonardo Fogassi, and Giacomo Rizzolatti. 1996. 'Action Recognition in the Premotor Cortex'. *Brain* 119 (2). Oxford Univ Press: 593–609.

Garey, Michael R, and David S Johnson. 1979. *Computers and Intractability*. Vol. 174. Freeman New York.

Gauch, Hugh G. 2003. *Scientific Method in Practice*. Cambridge University Press.

Gazzaniga, Michael S. 2005. 'Forty-Five Years of Split-Brain Research and Still Going Strong'. *Nature Reviews Neuroscience* 6 (8). Nature Publishing Group: 653–59.

Gervais, Raoul, and Erik Weber. 2012. 'Plausibility versus Richness in Mechanistic Models'. *Philosophical Psychology* 5089 (January): 1–14. doi:10.1080/09515089.2011.633693.

Gibbs, Raymond W., and Guy Van Orden. 2012. 'Pragmatic Choice in Conversation'. *Topics in Cognitive Science* 4 (1): 7–20. doi:10.1111/j.1756-8765.2011.01172.x.

Gigerenzer, G. 2008. 'Why Heuristics Work'. *Perspectives on Psychological Science* 3 (1): 20–29. doi:10.2307/40212224.

Gigerenzer, Gerd, Ralph Hertwig, and Thorsten Pachur. 2011. *Heuristics. The Foundations of Adaptative Behavior*. doi:10.1093/acprof:oso/9780199744282.001.0001.

Goldman, Alvin I. 2006a. *Simulating Minds*. *Philosophical Books*. Vol. 49. doi:10.1111/j.1468-0149.2008.459_15.x.

———. 2006b. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Goldstine, Herman H. 1980. *The Computer from Pascal to von Neumann*. Princeton University Press.

Goldstine, Herman H, and Adele Goldstine. 1946. 'The Electronic Numerical Integrator and Computer (ENIAC)'. *Mathematical Tables and Other Aids to Computation* 2 (15). JSTOR: 97–110.

Goodall, Jane. 2010. *Through a Window: My Thirty Years with the Chimpanzees of Gombe*. Houghton Mifflin Harcourt.

Goodman, Nelson. 1943. 'On the Simplicity of Ideas'. *The Journal of Symbolic Logic* 8 (4). JSTOR: 107–21.

———. 1955. 'Axiomatic Measurement of Simplicity'. *The Journal of Philosophy* 52 (24). JSTOR: 709–22.

———. 1958. 'The Test of Simplicity'. *Science* 128 (3331). American Association for the Advancement of Science: 1064–69.

———. 1959. 'Recent Developments in the Theory of Simplicity'. *Philosophy and Phenomenological Research* 19 (4). JSTOR: 429–46.

Gordon, Robert M. 2009. 'Folk Psychology as Mental Simulation'. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/folkpsych-simulation/#BM3.

Griinwald, Peter. 2005. 'Introducing the Minimum Description Length Principle'. *Advances in Minimum Description Length: Theory and*

Applications. MIT Press, 3.

Grunwald, Peter D., and Paul M.B. Vitanyi. 2003. 'Kolmogorov Complexity and Information Theory'.

Hartmanis, Juris, and Richard E Stearns. 1965. 'On the Computational Complexity of Algorithms'. *Transactions of the American Mathematical Society* 117. JSTOR: 285–306.

Heal, Jane. 1996. 'Simulation, Theory, and Content'. *Theories of Theories of Mind*. Cambridge University Press Cambridge, 75–89.

Heyes, Cecilia M, and Chris D Frith. 2014. 'The Cultural Evolution of Mind Reading.' *Science* 344 (6190): 1243091. doi:10.1126/science.1243091.

Hoffmann, Roald, Vladimir I Minkin, and Barry K Carpenter. 1997. 'Ockham's Razor and Chemistry'. *International Journal for the Philosophy of Chemistry* 3: 3–28.

Horgan, Terence, and John Tienson. 1992. 'Cognitive Systems as Dynamical Systems'. *Topoi* 11 (1): 27–43. doi:10.1007/BF00768297.

Hume, David. 2012. *A Treatise of Human Nature*. Courier Corporation.

Hutto, Daniel D. 2015. 'Basic Social Cognition without Mindreading: Minding Minds without Attributing Contents'. *Synthese*, July. doi:10.1007/s11229-015-0831-0.

Ito, Masao. 2012. *The Cerebellum: Brain for an Implicit Self*. FT press.

Jackson, Frank. 1982. 'Epiphenomenal Qualia'. *The Philosophical Quarterly* 32 (127). JSTOR: 127–36.

Jamieson, Dale. 2009. 'What Do Animals Think'. *The Philosophy of Animal Minds*. Cambridge University Press New York, 15–34.

Jerison, Harry J. 1973. 'Evolution of the Brain and Intelligence'. Academic Press, New York.

———. 1991. *Brain Size and the Evolution of Mind.* American Museum of natural history.

Kemeny, John G. 1955. 'Two Measures of Complexity'. *The Journal of*

Philosophy 52 (24). JSTOR: 722–33.

Kistermann, Friedrich W. 1998. 'Blaise Pascal's Adding Machine: New Findings and Conclusions'. *IEEE Annals of the History of Computing* 20 (1). IEEE: 69–76.

Kolmogorov, Andrei N. 1965. 'Three Approaches to the Quantitative Definition Ofinformation''. *Problems of Information Transmission* 1 (1): 1–7.

Kwisthout, Johan. 2012. 'Relevancy in Problem Solving: A Computational Framework'. *The Journal of Problem Solving* 5 (1): 4.

Levine, Joseph. 1983. 'Materialism and Qualia: The Explanatory Gap'. *Pacific Philosophical Quarterly* 64 (4): 354–61.

Li, Ming, and Paul Vitányi. 2009. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Science & Business Media.

Liebeskind, Benjamin J, David M Hillis, and Harold H Zakon. 2011. 'Evolution of Sodium Channels Predates the Origin of Nervous Systems in Animals'. *Proceedings of the National Academy of Sciences* 108 (22). National Acad Sciences: 9154–59.

Liggins, Martin E, David David Lee Hall, and James Llinas. 2008. *Handbook of Multisensor Data Fusion: Theory and Practice*. CRC press.

Lloyd, Seth. n.d. 'Measures of Complexity a Non-Exhaustive List'. http://web.mit.edu/esd.83/www/notebook/Complexity.PDF.

Löfgren, Lars. 1973. 'On the Formalization of Learning and Evolution'. In *Logic, Methodology and the Philosophy of Science IV.(Proceedings of the Fourth International Congress for Logic, Methodology and the Philosphy of Science, Bucharest, 1971)(Eds: Suppes, P et al)(Series Eds: Heyting, et Al. Studies in Logic and the Foundati*.

———. 1977. 'COMPLEXITY OF DESCRIPTIONS OF SYSTEMS: A FOUNDATIONAL STUDY'. *International Journal of General Systems* 3 (4): 197–214. doi:10.1080/03081077708934766.

Mach, Ernst. 1914. *The Analysis of Sensations, and the Relation of the Physical to*

the Psychical. Open Court Publishing Company.

Maguire, P, Philippe Moser, R Maguire, and Virgil Griffith. 2014. 'Is Consciousness Computable? Quantifying Integrated Information Using Algorithmic Information Theory'. *arXiv Preprint arXiv:1405.0126*. http://arxiv.org/abs/1405.0126.

Malik, Kenan. 2002. 'Man, Beast, and Zombie What Science Can and Cannot Tell Us About Human Nature'.

Mameli, Matteo. 2001. 'Mindreading, Mindshaping, and Evolution', 597–628.

Marois, Rene René, and Jason Ivanoff. 2005. 'Capacity Limits of Information Processing in the Brain.' *Trends in Cognitive Sciences* 9 (6). Elsevier: 296–305. doi:10.1016/j.tics.2005.04.010.

Marr, David. 1982. *Vision*. San Francisco: W.H. Freeman and Company.

Martignon, Laura, and Ulrich Hoffrage. 2002. 'Fast, Frugal, and Fit: Simple Heuristics for Paired Comparison'. *Theory and Decision* 52 (1): 29–71. doi:10.1023/A:1015516217425.

Massaro, Dominic W, and Nelson Cowan. 1993. 'Information Processing Models: Microscopes of the Mind'. *Annual Review of Psychology* 44 (1). Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA: 383–425.

McCulloch, Warren S, and Walter Pitts. 1943. 'A Logical Calculus of the Ideas Immanent in Nervous Activity'. *The Bulletin of Mathematical Biophysics* 5 (4). Springer: 115–33.

McGinn, Colin. 1999. *The Mysterious Flame : Conscious Minds in a Material World*. *Philosophical Review*. Vol. 110. Basic Books. doi:10.1215/00318108-110-2-300.

McLaughlin, Brian P. 2003. 'Computationalism, Connectionism, and the Philosophy of Mind'. In *The Blackwell Guide to the Phulosophy of Computing and Information*. Cambridge, UK: Blackwell Publishing Ltd.

Milkowski, Marcel. 2013. *Explaining the Computational Mind*. MIT Press

Cambridge, MA. doi:10.1073/pnas.0703993104.

Moore, G. 1965. 'Cramming More Components onto Integrated Circuits'. *Electronics* 38: 114–17.

Morin, Edgar. 1992. 'From the Concept of System to the Paradigm of Complexity Introduction: Mastering the Concept of System'. *Jownal of Social and E~oi~~i~~~ Systems* 15 (4): 371–85.

———. 2007. 'Restricted Complexity, General Complexity'. In *Worldviews, Science and Us*, 5–29. WORLD SCIENTIFIC. doi:10.1142/9789812707420_0002.

Moroz, Leonid L. 2009. 'On the Independent Origins of Complex Brains and Neurons'. *Brain, Behavior and Evolution* 74 (3). Karger Publishers: 177–90.

Nannini, Sandro. 2007. *Naturalismo Cognitivo. per Una Teoria Materialistica Della Mente*. Macerata: Quodlibet.

Neher, Erwin, and Bert Sakmann. 1976. 'Noise Analysis of Drug Induced Voltage Clamp Currents in Denervated Frog Muscle Fibres.' *The Journal of Physiology* 258 (3). Blackwell Publishing: 705.

Neisser, Ulric. 1967. *Cognitive Psychology*. Appleton Century Crofts.

Nietzsche, Friedrich. 2011. *The Will to Power*. Vintage.

Nordh, Gustav, and Bruno Zanuttini. 2005. 'Propositional Abduction Is Almost Always Hard'. In *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 19:534. LAWRENCE ERLBAUM ASSOCIATES LTD.

O'keefe, John, and Lynn Nadel. 1978. *The Hippocampus as a Cognitive Map*. Oxford University Press, USA.

Palmer, Stephen E. 1999. *Vision Science: Photons to Phenomenology*. MIT press.

Papadimitriou, Christos. 2014. 'Algorithms, Complexity, and the Sciences'. *Proceedings of the National Academy of Sciences* 111 (45): 15881–87. doi:10.1073/pnas.1416954111.

Pashler, Harold. 1994. 'Dual-Task Interference in Simple Tasks: Data and

Theory.' *Psychological Bulletin* 116 (2). American Psychological Association: 220.

Peirce, Charles Sanders. 1931. 'Collected Papers of Charles Sanders Peirce, Ed. C. Hartshorne and P. Weiss'. *Harvard University Press*. C. Hartshorne & P. Weiss). Harvard University Press, Cambridge, Mass.

Penrose, Roger. 1999. 'Precis of the Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics'. *Behavioral and Brain Sciences* 13 (4). Oxford University Press: 643–55.

Perconti, Pietro. 2008. 'L'autocoscienza'. *Cosa È, Come Funziona, a Cosa Serve, Laterza, Roma-Bari*.

———. 2015. *La Prova Del Budino: Il Senso Comune E La Nuova Scienza Della Mente*. Mondadori Università.

Perez-Zapata, Daniel, Virginia Slaughter, and Julie D Henry. 2016. 'Cultural Effects on Mindreading'. *Cognition* 146. Elsevier: 410–14.

Petersen, Steven E, Hanneke Van Mier, Julie A Fiez, and Marcus E Raichle. 1998. 'The Effects of Practice on the Functional Anatomy of Task Performance'. *Proceedings of the National Academy of Sciences* 95 (3). National Acad Sciences: 853–60.

Petitot, Jean. 1992. 'Physique Du Sens'. *Paris: Editions Du CNRS*.

———. 1999. *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford University Press.

Piccinini, Gualtiero. 2006. 'Computation without Representation'. *Philosophical Studies* 137 (2): 205–41. doi:10.1007/s11098-005-5385-4.

———. 2007. 'Computing Mechanisms'. *Philosophy of Science* 74 (4): 501–26. doi:10.1086/522851.

———. 2009. 'Computationalism in the Philosophy of Mind'. *Philosophy Compass* 4 (3): 515–32. doi:10.1111/j.1747-9991.2009.00215.x.

———. 2010a. 'Computation in Physical Systems'. Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.).

http://plato.stanford.edu/archives/fall2010/entries/computation-physicalsystems/.

———. 2010b. 'The Resilience of Computationalism'. *Philosophy of Science* 77 (5). JSTOR: 852–61.

———. 2015. 'Physical Computation: A Mechanistic Account'.

Piccinini, Gualtiero, and Sonya Bahar. 2013. 'Neural Computation and the Computational Theory of Cognition'. *Cognitive Science* 37 (3): 453–88. doi:10.1111/cogs.12012.

Piccinini, Gualtiero, and Andrea Scarantino. 2010. 'Computation vs. Information Processing: Why Their Difference Matters to Cognitive Science'. *Studies in History and Philosophy of Science Part A* 41 (3): 237–46. doi:10.1016/j.shpsa.2010.07.012.

Plum, F. 1991. 'Coma and Related Global Disturbance of the Human Conscious State'. In *Peters A. and Jones E. G. (Eds), Normal and Altered States of Function*, edited by A. Peters and E.G. Jones. New York: Plenum Press.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. Routledge.

Putnam, H. 1960. 'Mind and Machines'. In *Mind, Language and Reality*, 362–85.

———. 1967. 'The Nature of Mental States'. In *Mind, Language and Reality*, 429–40.

———. 1975. *Mind, Language and Reality*. Cambridge, UK: Cambridge University Press.

Randel, Nadine, and Gaspar Jekely. 2016. 'Phototaxis and the Origin of Visual Eyes'. *Phil. Trans. R. Soc. B* 371 (1685). The Royal Society: 20150042.

Reif, John H, and Zheng Sun. 2003. 'On Frictional Mechanical Systems and Their Computational Power'. *SIAM Journal on Computing* 32 (6). SIAM: 1449–74.

Rissanen, Jorma. 1978. 'Modeling by Shortest Data Description'. *Automatica* 14 (5). Elsevier: 465–71.

Rizzolatti, Giacomo, and Corrado Sinigaglia. 2008. *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. Oxford University Press, USA.

Roth, Gerhard. 2012. 'Is the Human Brain Unique?' In *The Theory of Evolution and Its Impact*, 175–87. Springer.

Roth, Gerhard, and Ursula Dicke. 2013. 'Evolution of Nervous Systems and Brains'. In *Neurosciences-From Molecule to Behavior: A University Textbook*, 19–45. Springer.

Rumelhart, D.E., and J.L. McClelland. 1986. *Paralle Distributed Processing: Exploration in the Microstructure of Cognition*. Cambridge, MA: MIT Press.

Ryan, Tomás J, and Seth G N Grant. 2009. 'The Origin and Evolution of Synapses'. *Nature Reviews Neuroscience* 10 (10). Nature Publishing Group: 701–12.

Saltzman, Elliot L. 1995. 'Dynamics and Coordinate Systems in Skilled Sensorimotor Activity'. *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press Cambridge, MA, 149–73.

Searle, John R. 1990. 'Is the Brain a Digital Computer?' In *Proceedings and Addresses of the American Philosophical Association*, 64:21–37. JSTOR.

———. 1992. *The Rediscovery of the Mind*. MIT press.

Searle, John R. 1980. 'Minds, Brains, and Programs'. Edited by John Haugeland. *Behavioral and Brain Sciences* 3 (3). Cambridge Univ Press: 417. doi:10.1017/S0140525X00005756.

Sengpiel, Frank. 1997. 'Binocular Rivalry: Ambiguities Resolved'. *Current Biology* 7 (7). Elsevier: R447–50.

Shannon, C E. 1948. 'A Mathematical Theory of Communication'. Edited by Pier Luigi Luisi and Pasquale Stano. *Bell System Technical Journal*, The mathematical theory of communication, 27 (July 1928). ACM: 379–423. http://dl.acm.org/citation.cfm?id=584093.

Simon, Herbert A. 1990. 'Invariants of Human Behavior'. *Annual Review of Psychology* 41 (1). Annual Reviews 4139 El Camino Way, PO Box 10139,

Palo Alto, CA 94303-0139, USA: 1–20.

Sober, Elliott. 1990. 'Explanation in Biology: Let's Razor Ockham's Razor'. *Royal Institute of Philosophy Supplement* 27. Cambridge Univ Press: 73–93.

———. 2002. 'What Is the Problem of Simplicity?' *Simplicity Inference and Econometric Modelling*, 13–32. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.75.454.

Solomonoff, Ray J. 1964. 'A Formal Theory of Inductive Inference. Part I'. *Information and Control* 7 (1). Elsevier: 1–22.

Sperry, Roger W. 1976. 'Mental Phenomena as Causal Determinants in Brain Function'. In *Consciousness and the Brain*, 163–77. Springer.

Steinhart, Eric. 2002. 'Logically Possible Machines'. *Minds and Machines* 12 (2). Springer: 259–80.

Stich, Stephen P. 1979. 'Do Animals Have Beliefs?' *Australasian Journal of Philosophy* 57 (1). Taylor & Francis: 15–28.

Stich, Stephen, and Ian Ravenscroft. 1994. 'What Is Folk Psychology?' *Cognition* 50 (1–3). Elsevier: 447–68. doi:10.1016/0010-0277(94)90040-X.

Swade, Doron. 2011. 'Pre-Electronic Computing'. In *Dependable and Historic Computing*, 58–83. Springer.

Swade, Doron D. 2005. 'The Construction of Charles Babbage's Difference Engine No. 2'. *IEEE Annals of the History of Computing* 27 (3): 70–78.

Swinburne, Richard. 2001. 'Epistemic Justification'.

Thagard, Paul. 1993. 'Computational Tractability and Conceptual Coherence: Why Do Computer Scientists Believe That P≠ NP?' *Canadian Journal of Philosophy* 23 (3). Taylor & Francis: 349–63.

Thagard, Paul, and Karsten Verbeurgt. 1998. 'Coherence as Constraint Satisfaction'. *Cognitive Science* 22 (1). Wiley Online Library: 1–24.

Thelen, Esther. 1995. 'Time-Scale Dynamics and the Development of an Embodied Cognition'. *Mind as Motion: Explorations in the Dynamics of Cognition*. Massachusetts Institute of Technology, Cambridge, MA, 69–

100.

Thom, René. 1988. ‘Esquisse D’une Sémiophysique. Physique Aristotéliciene et Théorie Des Catastrophes’. Interéditions, Paris.

Tononi, Giulio. 1998. ‘Consciousness and Complexity’. *Science* 282 (5395): 1846–51. doi:10.1126/science.282.5395.1846.

———. 2003. *Galileo E Il Fotodiodo: Cervello, Complessità E Coscienza*. Bari: Laterza.

———. 2004a. ‘An Information Integration Theory of Consciousness’. *BMC Neuroscience* 5 (1). BioMed Central: 42. doi:10.1186/1471-2202-5-42.

———. 2004b. ‘An Information Integration Theory of Consciousness.’ *BMC Neuroscience* 5 (November): 42. doi:10.1186/1471-2202-5-42.

———. 2008. ‘Consciousness as Integrated Information: A Provisional Manifesto.’ *The Biological Bulletin* 215 (3). MBL: 216–42. http://www.ncbi.nlm.nih.gov/pubmed/19098144.

———. 2010. ‘Information Integration: Its Relevance to Brain Function and Consciousness.’ *Archives Italiennes de Biologie* 148 (3): 299–322. http://www.ncbi.nlm.nih.gov/pubmed/21175016.

———. 2012. *Phi: A Voyage from the Brain to the Soul*. Random House Digital, Inc.

Tononi, Giulio, and David Balduzzi. 2009. ‘Toward a Theory of Consciousness’. In *The Cognitive Neurosciences*, edited by Michael S Gazzaniga, 54:1201–17. The MIT Press. doi:10.1006/ccog.1999.0390.

Tononi, Giulio, G M Edelman, and O Sporns. 1998. ‘Complexity and Coherency: Integrating Information in the Brain.’ *Trends in Cognitive Sciences* 2 (12): 474–84. http://www.ncbi.nlm.nih.gov/pubmed/21227298.

Tononi, Giulio, and Christof Koch. 2008. ‘The Neural Correlates of Consciousness: An Update.’ *Annals of the New York Academy of Sciences* 1124 (March): 239–61. doi:10.1196/annals.1440.004.

Townsend, James T, and Jerome Busemeyer. 1995. ‘Dynamic Representation

of Decision-Making'. *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press, 101–20.

Trautteur, G. 1999. 'Analog Computation and the Continuum-Discrete Conundrum'. *R. Lupacchini E G. Tamburrini (a Cura Di), Grounding Effective Processes in Empirical Laws. Reflections on the Notion of Algorithm, Bulzoni, Roma*.

Tsotos, John K. 1990. 'Analyzing Vision at the Complexity Level'. *Behavioral and Brain Sciences* Behavioral (13): 423–69.

Turing, A M. 1936. 'On Computable Numbers, with an Application to the Entscheidungs Problem'. *Proceedings of the London Mathematical Society*, 2, 42 (42). London Mathematical Society: 230–65. doi:10.1112/plms/s2-42.1.23.

Valiant, Leslie. 2013. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books.

van Gelder, T. 1998. 'The Dynamical Hypothesis in Cognitive Science.' *The Behavioral and Brain Sciences* 21 (5): 615-628-665. doi:10.1017/S0140525X98001733.

Van Gelder, Tim. 1995a. 'What Might Cognition Be, If Not Computation?' Edited by William G Lycan. *Journal of Philosophy* 92 (7). Blackwell Publishers: 345–81. doi:10.2307/2941061.

———. 1995b. 'What Might Cognition Be If Not Computation?' *The Journal of Phylosophy* 92 (7): 345–81.

Van Gelder, Tim, and Robert Port. 1998. *Mind as Motion*. *Exploration in the Dynamics of Cognition*. MIT Press Cambridge, MA. doi:10.1007/s13398-014-0173-7.2.

van Rooij, Iris. 2008. 'The Tractable Cognition Thesis'. *Cognitive Science: A Multidisciplinary Journal* 32 (6): 939–84. doi:10.1080/03640210801897856.

Van Rooij, Iris. 2008. 'The Tractable Cognition Thesis'. *Cognitive Science* 32 (6). Psychology Press: 939–84. doi:10.1080/03640210801897856.

van Rooij, Iris, and T. Wareham. 2007. 'Parameterized Complexity in Cognitive Modeling: Foundations, Applications and Opportunities'. *The Computer Journal* 51 (3): 385–404. doi:10.1093/comjnl/bxm034.

van Rooij, Iris, and Todd Wareham. 2012. 'Intractability and Approximation of Optimization Theories of Cognition'. *Journal of Mathematical Psychology* 56 (4): 232–47.

Van Rooij, Iris, Cory D. Wright, and Todd Wareham. 2010. 'Intractability and the Use of Heuristics in Psychological Explanations'. *Synthese* 187 (2): 471–87. doi:10.1007/s11229-010-9847-7.

Vitanyi, Paul. 1998. 'Simplicity , Information , Kolmogorov Complexity , and Prediction 1 Introduction', 1–23.

Vitányi, Paul, and Ming Li. 2000. 'Minimum Description Length Induction, Bayesianism, and Kolmogorov Complexity'. *Information Theory, IEEE Transactions on* 46 (2). IEEE: 446–64.

Wareham, Todd, Johan Kwisthout, Pim Haselager, and Iris Van Rooij. 2011. 'Ignorance Is Bliss: A Complexity Perspective on Adapting Reactive Architectures'. *2011 IEEE International Conference on Development and Learning ICDL*. IEEE. http://ieeexplore.ieee.org/ielx5/6031618/6037311/06037337.pdf?tp=&arnumber=6037337&isnumber=6037311.

Weaver, Warren. 1948. 'Science and Complexity'. *American Scientist* 36 (4): 536–44. doi:yes.

Wegner, Peter, and Dina Goldin. 2003. 'Computation beyond Turing Machines'. *Communications of the ACM* 46 (4). ACM: 100. doi:10.1145/641205.641235.

Wimsatt, William C. 1997. 'Aggregativity: Reductive Heuristics for Finding Emergence'. *Philosophy of Science* 64 (S1): S372. doi:10.1086/392615.

Wright, John. 1991. *Science and the Theory of Rationality*. Avebury Aldershot.

Yamashita, Yuichi, and Jun Tani. 2008. 'Emergence of Functional Hierarchy in

a Multiple Timescale Neural Network Model: A Humanoid Robot Experiment.' *PLoS Computational Biology* 4 (11): e1000220. doi:10.1371/journal.pcbi.1000220.

Zakon, Harold H. 2012. 'Adaptive Evolution of Voltage-Gated Sodium Channels: The First 800 Million Years'. *Proceedings of the National Academy of Sciences* 109 (Supplement 1). National Acad Sciences: 10619–25.

Zawidzki, Tad. 2009. 'Theory of Mind, Computational Tractability, and Mind Shaping: 2009 Performance Metrics for Intelligent Systems Workshop'. In *Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*, 149–54. ACM.

Zawidzki, Tadeusz W. 2008. 'The Function of Folk Psychology: Mind Reading or Mind Shaping?' *Philosophical Explorations* 11 (3). Taylor & Francis: 193–210.

Zawidzki, Tadeusz W. 2008. 'The Function of Folk Psychology: Mind Reading or Mind Shaping?' *Philosophical Explorations* 11 (3): 193–210. doi:10.1080/13869790802239235.

Zawidzki, Tadeusz Wieslaw. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. MIT Press.

Zigmond, Michael J, and Floyd E Bloom. 1999. 'Fundamental Neuroscience'.

Zuse, Konrad. 1982. 'The Outline of a Computer Development from Mechanics to Electronics'. In *The Origins of Digital Computers*, 175–90. Springer.