MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Scientific Representations in Computational Views of Cognitive Science“

verfasst von / submitted by

Raphael Gustavo Aybar Valdivia

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019/ Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Joint Degree Programme MEi: CogSci Cognitive Science UG2002
Betreut von / Supervisor:	Dr. -Ing. Igor Farkas, PhD

Abstract

The notion of representation is at the heart of cognitive science's enterprise, which regards it mainly in terms of internal or mental state. Recent works in the philosophy of science have reconsidered these views of representations, suggesting that they are also elements of the scientific frameworks that support the discovery and understanding of cognitive phenomena. Scientific representation refers to the type of representation operating in the scientific inquiry that is currently conducted using models. Models, in turn, have various epistemic possibilities which are not restricted to representation. It is argued that only after de-idealizations, additions of constraints and justifications of the assumptions, models can serve for representing phenomena.

Considering the manifold of possibilities of models, this research asks the following question: how scientific representations are achieved in cognitive science? This project reviews the main recent literature on scientific representation from philosophy of science and phenomenological approaches to the act of representing in order to sketch a general interpretation of scientific representation. After this, it reviews the processes of idealization, de-idealization, analogical reasoning, addition of constraints and explicitation of assumptions involved in the construction of models intended to represent. These considerations are supported by various examples of the use of computational models in cognitive science that clarify how these theoretical approaches fit some current scientific practices. It also discusses how two of the main paradigms in the field (cognitivism and connectionism) understand representations.

The thesis puts forth a philosophical argument which is relevant for the epistemology of cognitive sciences. Its strategy is clearly interdisciplinary as far as it studies a type of modeling employed in different areas of cognitive science.

Table of Contents

1. Introduction

The concept of representation seems to be at the heart of cognitive science. It is assumed that representations play a central role in cognitive processes (at least in higher cognitive domains). Representations are frequently considered as the contents of mental states, which are assumed to be the object that the sciences of the mind intend to comprehend (Shea, 2018). The central hypothesis of cognitive science: “thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures” (Thagard, 2005, p. 10), inspires almost all current cognitive science research. Representations have also been examined in theories of mind (ToM) in terms of states such as beliefs, desires, thoughts, etc., which are intentional or about something. These representations might designate or target the content of the intentional relation (qualia), and they could be modal, bimodal, multimodal, supramodal or amodally apprehended (Barsalou, Santos, Simmons, Wilson, 2008; Gallese and Lakoff, 2005).

Recent research in cognitive psychology and cognitive linguistics are concerned to the debate of the modality of representations and concepts, while research in artificial intelligence and cognitive neuroscience currently employ notions of representation for describing the behavior of cognitivist models, connectionists models, etc. In this context, the notion of structural representation, a relation of resemblance or isomorphism between scientific models and phenomena, has been presented as an adequate account to explain how representations work in scientific domains. More precisely, a correspondence between mental states and contents in the brain is expected.

Authors such as Morgan (2014) claim than intentional accounts and theories of mind are nowadays outdated in cognitive science. Nonetheless, neuroscientific research (Schurz and Perner, 2015), brain injury studies (Dennis et al., 2013), and studies of the Autism Spectrum Disorder (Fletcher–Watson et al., 2014) still employ them. Despite its detractors, ToM is still a valid approach of cognitive processes, although its extent and assumptions are not well–defined.

The concept of representation remains in a similar situation. Although there are current investigations on the nature of representation in cognitive science (see, for instance, Shea, 2018; Gómez–Ramirez, 2014), they are mainly concerned with the nature of mental representations without realizing the role they play in scientific research. In this respect, there exist research on scientific representation in current philosophy of science, but it has not been considered in cognitive science.

Considering this lack, the following research develops a different approach to representations in the field of cognitive science based on a basic distinction between two senses of the term: representations, on the one hand, can be regarded as an object of study in cognitive processes; on the other hand, they can be treated as elements of the scientific framework by means of which cognitive scientists make sense of the mental (theoretical attitude). The means by which scientists study the mind must be distinguished from what is studied. Certainly, the extent of the instruments used by scientists to study mental processes go beyond representational artifacts – there are as well non–representational techniques in this discipline. This distinction is meant to make salient certain qualities of representations: they are elements of cognitive processes and tools that scientists use to understand phenomena. It cannot simply be argued, as radical enactivists do, that cognition does not involve representation if the very basis of our scientific knowledge employs representations in various ways. Further, if representations are part of higher cognitive domains, they must be part as well of scientific cognition.

Whereas most research in cognitive science and philosophy of mind refers only to the first meaning of representation (mental states), the strategy developed in this thesis differs from these perspectives as far as it is concerned with the nature of the representations employed in scientific research. In order to do this, this research takes into account various philosophical debates from the early nineties of the last century. These debates were concerned to the nature of scientific representations and later had evolved to a more profound discussion of the use of models in sciences. Additionally, the research reviews the phenomenology of theoretical acts developed by Husserl because it confronts some non–explored questions in these debates.

The attempt of looking at science as a particular type of cognition underlies both accounts of representation. When cognitive scientists approach subjects such as representations, perception, or other cognitive processes, they are using different specific cognitive processes at the same time. Since the line between representations that are part of mental processes and representations that are also part of scientific frameworks is diffuse, it is a pendant task to create a taxonomy that classifies representations by considering their nature and function.

Considering this, this research asks the following questions:

- Which are the elements involved in scientific representation?
- Under which circumstances do computational models serve to represent cognitive phenomena?

The first question is answered by discussing different philosophical approaches to scientific representation and models (semantic, pragmatic and phenomenological). The main answer provided to this question is that scientific representation is a relation produced in the interaction between skilled agents that intend to use models to stand for things. Models are not inherently representational. To achieve representation, modelers need to add certain constraints and make assumptions by considering the current knowledge of the phenomena they are interested in (for example, the existence of mechanisms that underlie their behaviors). An answer to the first question is needed to examine the next one. The second question intends to use these philosophical discussions on scientific representations and models to make sense of how they are used by cognitive scientists.

Computational modeling has been chosen as the phenomenon this research examines because it is one of the main representational techniques employed by cognitive scientists. Computational models are used by cognitive scientists that assume that minds are either computing machines or that the analogy between both is fruitful for analyzing cognitive processes. In this regard, scientists commit with various assumptions regarding the computational nature of mental processes. Since they also assume that representations are part of cognitive processes, they become theoretical entities that underlie computational models of the mind. In this perspective, ToM, the computational theory of mind, and the computational–representational understanding of the mind can be regarded as specific theoretical stances adopted by cognitive scientists in the attempt to understand cognitive processes. This means they are different interpretations of cognitive phenomena that postulate various assumptions about the mind to gain knowledge or other practical concerns.

Furthermore, this research explores some representational strategies in cognitive modeling by considering how artificial neural network models are used to represent cognitive phenomena. In what follows, a more detailed review of the contents of each section of the thesis is presented.

Section 2.1 discusses various accounts of scientific representation developed by recent philosophers of science (semantic views). Using these and other insights this section introduces the different elements that compose scientific representation – a relation involving models, theories and phenomena, focusing on how cognitive science uses computational models to stand for cognitive processes. According to these philosophers of science, representation is a relation composed by a source (vehicle of representation) and a target system (Suárez, 2003b). It is presumed that sources and target systems share relations of similarity or isomorphism.

Sections 2.1 and 2.2 review the inconsistencies of the semantic view, suggesting that scientific representations should be rather analyzed from pragmatic and phenomenological views. These two alternative views have in common an emphasis on agents as essential components of scientific representation. According to the pragmatic views, it is a non–sense to postulate a subject–independent quality of similarity or isomorphism. These properties are rather proposed by skilled agents in certain contexts of inquiry for achieving certain goals that could be representational (although they are not restricted to this possibility). A more detailed explanation of these arguments is offered in the introduction of section 2.

In section 2.2, some phenomenological concepts such as natural–scientific/theoretical attitude and theoretical acts are reviewed in order to formulate the concept of theoretical stance. According to Husserl, scientific attitudes and acts determine the phenomena they are interested in by postulating both their psycho–physical existence and the existence of certain structures that underlie their behaviors. Following this, scientific activities should not be reduced to specific outcomes of scientific knowledge. The rise of data sciences, for instance, has had an impact on the commonsense views about the importance of prediction in science. Many scientists nowadays believe that prediction is from far the most important validation criterion for scientific knowledge. However, predictive power could come at the cost of reducing the representational power of a certain model. Not all predictive models reproduce the minimal conditions (causal) that can explain how certain phenomena arise (see section 3.2). In that sense, the outcome (predictive power) can be distinguished from the broader extent of scientific validation.

As stated, pragmatic and phenomenological views emphasize the role agents play in representational relations. Without *agents that intend to represent*, scientific representations are not possible, while the expertise and skill of these agents are needed to make models powerful representational devices. These approaches stress that representational power is not inherent to the model and rather it is something that must be stated and achieved in the process of constructing models in which assumptions are taken and constraints are added.

Although this research is a theoretical exploration of the uses of scientific representation, the phenomena that guide its analyses are the uses of computational models in cognitive science. These practices have not been chosen by coincidence. The interest in exploring them respond to a consideration of the modeling strategies employed in cognitive science. Cognitive science has been very influenced by the development of computer sciences from its origins, and this influence – despite important changes – persists until now. Various computational models are being used for modeling strategies of cognitive phenomena. But this influence is not restricted to pragmatically–oriented approaches to modeling. The use of computational models is accompanied and supported with theoretical standpoints that postulate the use of models as theoretical devices for representing cognitive processes. Section 3 explores these theoretical possibilities in more details.

Cognitive scientists assume that cognitive processes are either computations or at least involve some sort of computing. Certainly, this hypothesis has many detractors such as autopoietic enactivists who assume that cognition is rather an exclusive property of living organisms. Despite these claims, the analogy between cognition and computation is sufficiently strong to be considered one of the main assumptions in the sciences of the mind. In fact, two of the main paradigms in cognitive science – cognitivism and connectionism – subscribe to the computational view of the mind. Their theoretical bases are the computational theory of mind (cognitivism) and the computational–representational understanding of the mind (connectionism).

Section 3.1 introduces these different computational views and discusses how representations are understood by them. The main idea is that whereas cognitivist approaches and computational theories of mind endorse the metaphysical hypothesis that cognition is computation, connectionism and the computational–representational understanding of the mind, instead, can be related to pragmatic approaches to modeling that do not consider that representations are the only sources of scientific knowledge. The general idea is that the analogy between computation and cognition is useful for gaining knowledge of the mind (not necessarily via representing), and it can serve to other practical concerns as well. Representation in this context is only achieved if certain construction assumptions are adopted and specific constraints are added to artificial neural network models. The assumptions and constraints needed for representing are explored in section 3.2.

Although artificial neural network models are inspired by real connections occurring in the brain, this inspiration is not enough for representing. These models are highly abstract by themselves, and without certain adding certain constraints they remain as mere computational templates or black–box units. These constraints are added in the process of building models, which involves different sort of techniques. Firstly, an analogy between better–known and lesser–known domains need to be established, otherwise, a model would not be directed at any phenomena. Computational models are better–known controlled domains. They can be used for modeling phenomena and even when direct experimentation is not possible.

When an analogy between two domains is introduced, assumptions regarding the nature of the phenomena must be stated. These assumptions serve to specify the ontology implied by these models. This step is necessary because it makes explicit what is the extent of the analogy between the two domains. Based on their knowledge about cognitive phenomena, modelers postulate the existence of mechanisms that underlie these phenomena. They use computational models to track these mechanisms. Thus, the existence of these mechanisms and relations of similarity or isomorphism between models and these entities should be justified. If assumptions are not stated, there is no way to justify that certain model represents a certain phenomenon.

Once these assumptions are stated, constraints must be added into a model. Artificial neural networks are by themselves too abstract for representing anything in the world. Without adding certain constraints, they may serve for practical purposes – for instance, to gain predictive power – but they might not be useful for representing phenomena. These constraints are added by considering the features of the phenomena they are directed at.

Finally, although idealization could be regarded as a reductive strategy in computational modeling, it is one of the main representational techniques needed for artificial neural network models. Idealization is required for various purposes:

phenomena are sometimes non–tractable and scientists need to create simplified models for gaining computational tractability; phenomena are complex and minimal models are indispensable for delineating the basic mechanisms that could give a causal explanation of their behaviors. Finally, sometimes one single model is not enough for giving an account of phenomena. As a result, the use of various models is required. Of course, not all types of idealization can serve for representing, and these strategies need to be complemented with de–idealizations that make them less abstract and closer to the phenomena they are directed at.

Summing up, this research examines how computational models can be used for representing cognitive phenomena. It argues that only under certain specific conditions agents can use computational models to stand for things. Next sections provide an understanding of scientific representations by exploring and discussing the conditions and assumptions that make scientific models useful for representing phenomena.

The main expected finding of this research is an understanding of the general different senses of representations in cognitive science. Other insights include arguments against semantic views of scientific representation and in favor of pragmatic and phenomenological views of scientific representation that could provide philosophical groundings to the use of computational modeling despite the enactive criticism.

Finally, this thesis attempts to develop a philosophical argument which is relevant for the epistemology of cognitive science. This strategy is clearly interdisciplinary as far as computational models are used in various disciplines that investigate cognitive processes.

2. Philosophical approaches to Scientific Representation

Science is a type of knowledge and, as such, it is not radically different from other forms of knowledge such as experience, belief, conjecture, or faith. However, a distinctive trait of scientific knowledge is the methodical procedure by means of which knowledge is obtained. While faith, for instance, does not need to be tested – at least in the sense of supporting or rejecting some hypotheses, scientific knowledge does need this. The same goes for experiential knowledge that relies more upon the habits and regularities that people find in nature and in the social world than in artificial or experimental ways to prove their beliefs.

A scientific method is a technique that serves to validate or reject a belief or a system of beliefs. Scientific knowledge is acquired and constructed by following a series of steps that are inter–subjectively developed and established by scientific communities. In science, for example, "(…) repeatability typically requires intersubjective agreement among scientists observing similar events at different times and in different geographical locations" (Velmans, 1999, p. 305). Scientists themselves use to follow methods rather than create them from the scratch, and they can also contribute to their improvement through criticism, either by revealing their limitations, loose ends, misunderstandings, non–explicit assumptions, prejudices, which are often carried out in their applications. In this way, they contribute to their optimization or, in the worst scenarios, to their rejection. In some special cases, scientists create new methods or radically change their practices.

Indeed, current scientific practices retain many elements inherited from tradition, such as formulating and testing hypotheses, research questions, and even the very definition of a research topic. However, many scientific practices have radically changed in the history. For instance, it is traditionally understood that science proposes theories, that is, explanatory systems of certain parts of reality. Without denying their value, modern science now also acquires knowledge by designing concrete or ideal entities – its ontological status is still a subject of debates – called models, which simulate in some respects their target realities.

Within the philosophical tradition, scientific knowledge (*epistêmê*) has been considered more firm and stable than opinion (*doxa*) (Szaif, 2007, p. 266), although it is far from clear what both share. Hence, it seems necessary to ask: what kind of similarities does scientific knowledge share with non–scientific cognition? Although there are thousands of research papers dedicated to the various forms of non–scientific cognition, exploring dimensions such as decision making, memory, perception, problem solving, etc. (see, for instance, Wang, 2007; Chubb, Dosher, Shiffrin and Zhong, 2013; Pecher and Zwaan, 2010; Newton, 2016), there is a lack of interest to find how they are manifested in scientific cognition. Nevertheless, in recent cognitive science and philosophy of mind there is a growing interest in understanding the continuity between cognition and action (Thompson, 2007). This is the perspective of the enactive approach, which considers that "cognition is not the representation of a pre–given world by a pre–given mind but it is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs" (Varela, Thompson and Rosch, 2016, p. 9). In this sense, enactivists propose continuity from lower to higher forms of cognition (Kiverstein and Rietveld, 2018).

Despite the intentions of its founders (Varela and Maturana), more recent proponents of the enactivism seem to be more interested in the low–level forms of cognition (e.g. Shapiro, 2007), and some of them hold anti–representational views of cognition (e.g. Chemero, 2009; Hutto and Myin, 2013; Fuchs and de Jaegher, 2009). When intending to explain higher forms of cognition such as social cognition or consciousness, they tend to replicate the kind of explanations used for lower levels to higher levels (e.g. de Jaegher and Di Paolo, 2007). This leads to abstract formulations and untestable claims that the scientific community does not take as serious alternatives to theories of mind (in social cognition) – despite being well appreciated in the philosophical community.

To sum up, high–level forms of cognition such as consciousness, mathematical and logical thinking, or even representations (and by extension scientific ones) are unsatisfactorily explained by the enactive approach (Stewart, Gapenne and Di Paolo, 2010, viii). However, if enactivism is on the right track when claiming that “there is a deep continuity in the principles of self–organization from the simplest living things to more complex cognitive beings” (Varela, Thompson & Rosch, 2016, xix), then scientific cognition – considered a highly complex and sophisticated form of cognition – should also be enacted, embodied, embedded and perhaps extended. Unfortunately, there is no single account of scientific cognition from the enactive point of view; thus, the specific way in which scientists ‘enact’ and commerce with the world constituting themselves and the world through these interactions simply remains as a mystery.

Fortunately, other philosophers have tried to explain what scientific representation consists of, as well as what is the nature of scientific acts. In what follows, analytic and phenomenological views of scientific representation will be introduced to determine what would be fair to expect from the manifold of notions of representation employed in cognitive science. This procedure will provide some relevant cues to the next sections, which will be dedicated to the theories of mind and the computational understanding of the mind as part of a representational strategy in the context of certain scientific activities.

Section 2.1 introduces semantic and pragmatic views of scientific representations and models. The first holds the idea that scientific models represent reality or at least its structure by maintaining relations of similarity and/or isomorphism with their target systems. The section discusses the weak points of semantic views, e.g., their inability to justify the possibility of misrepresentation. In opposition, pragmatic views of scientific representation introduce context dependencies and agents in the equation but understand the intentionality of representational acts in an impoverished minimal sense (as a mere directedness).

Section 2.2 argues that representational acts can be better understood from phenomenological perspectives. It introduces the phenomenology of theoretical acts developed by Husserl in several texts and proposes that an essential feature of scientific knowledge is to postulate the existence of its object. Additionally, it explains the mediated or derivative character of scientific acts, how they are validated, their intersubjective nature, and it suggests that they are in a continuum with lower–level forms of cognition. This section concludes with a summary listing the characteristics of scientific representation. Using these ideas, the next sections will assess scientific representations in the computational–representational understanding of the mind, the computational theory of mind, as well as in cognitivism and connectionism.

2.1. Semantic and pragmatic views of scientific representations

Traditionally, philosophers have considered science as “an activity aiming at representing part of the world” (Frigg 2006, p. 49). This activity is carried out using entities that stand for certain phenomena, or that in some way invoke them. Philosophers of science call these entities ‘sources’ or ‘vehicles’ of the representational relation (this section refers to them as sources). These sources correspond to what they designate as ‘scientific representation’, a concept that covers data sets, concepts, theories, and models. A representational relation is established between these sources and pretended target systems, i.e., the phenomena they are directed at. “The term ‘target system’ or simply ‘target’ has been used to refer to what is represented, such as a physical object, a process, a population or a phenomenon” (Knuuttila, 2011, p. 264).

Although these concepts are quietly well defined, philosophers of science do not agree on what supports the representational relation, in other words, in virtue of what the sources represent their target systems. Among the alternatives it is claimed that sources and targets are bond by axiomatic structures or sets of statements in certain language (syntactic views), non–linguistic models (semantic views), or activities of certain users (pragmatic views) (Suárez and Pero, 2019, p. 344). The recent debate in analytic philosophy of sciences takes into account the last two, while syntactic views are currently considered as erroneous[1]. Because of this, the following section only discusses semantic and pragmatic views.

The semantic view of representations and models "(…) is perhaps the only global analysis of science in these philosophically fractured, post–Kuhnian times" (French and Ladyman, 1999, p. 103). It confronts both the more skeptical viewpoints regarding the possibility to find something that characterizes any scientific activity as well as the naïve positions that posit the possibility of sciences in a formal logic that – in some unknown way – allows scientific representations to be directed at reality. Semantic views aim to explain what scientific knowledge consists of, what are its components, and in virtue of what they gain knowledge of the world.

The different versions of the semantic view have in common an understanding of scientific theories in terms of extra–linguistic entities rather than sets of propositions or axioms (Suárez & Pero, 2019, p. 349). This means that theories are neither logical structures, nor sets of statements, and usually incorporate the aside elements such as models and instruments. Furthermore, semantic views hold that principles of

[1] An explanation of why syntactic views is rejected is provided by Bailer–Jones (2009, pp. 127–128).

theories, axioms, and laws are not directed at phenomena. In other words, they "(…) cannot by themselves be used to make any direct claim about the world" (Giere, 2010, p. 270). Highly abstract components of theories need certain mediation to be able to refer and/or represent to concrete phenomena. This mediation, which constrains and specifies them, is an essential component of theories that makes them capable of referring to things in the world. Semantic views affirm that constructing models is a strategy to accomplish this mediation. Considering this, they assert that models, and not theories, can be used to represent phenomena. According to Frigg (2006), models

> (…) play an essential role in the acquisition and organization of scientific knowledge. We often study a model to discover features of the thing that it stands for. For instance, we study the nature of the hydrogen atom, the dynamics of populations, or the behavior of polymers by studying their respective models (p. 49).

Standing for something else is a defining characteristic of the modern sense of representing, which is close to the attempt of "(…) substituting something absent with something present" (Knuuttila, 2011, p. 263). Models substitute the thing they stand for because the study of their properties replaces the direct study of target systems. It is assumed that this is a valid procedure through which knowledge is gained (Frigg, 2006, p. 49). Nevertheless, models are not direct representations of phenomena as far as they can, for instance, simplify the phenomena or depict only a few properties that interest in a certain context of inquiry. In this sense, they are indirect representations, as Godfrey–Smith (2006) points out in the following:

> What is most distinctive of model–based science is a strategy of indirect representation of the world (...). The modeler's strategy is to gain

> understanding of a complex real–world system via an understanding of simpler, hypothetical system that resembles it in relevant respects (p. 726).

Apart from being substitutes, models are also simplifications of their target systems. It would be impossible that the study of a model informs something if they were too complex as them. Thus, "(…) only a few properties are attributed instead of striving to represent some real target systems directly" (Knuuttila, 2011, p. 266). In that sense, models do not need to be accurate in a strong sense. Consider the following example: the connectionist self–organizing map model (SOM) of language acquisition. This model could serve to comprehend the emergent properties of higher levels domains of human cognition such as lexical categories (Li and Zhao, 2013, p. 2). However, the difference between these models and actual human language acquisition are too salient since the model reduces language acquisition to processing of a few inputs that in no case correspond to all the influences humans have while learning a language (in this sense, this model is a simplification). Cognitive scientists, though, do not always pretend that models represent human or other species cognition as such. Some of them might argue that they only serve for depicting highly idealized or abstract forms of cognition for practical concerns.

Semantic views hold that similarities between the model and the target system make the first to stand for the latter. In the example, the SOM is an unsupervised learning model that "uses no explicit error signal to adjust the weights between input and output", "requires no explicit teacher", and in which "learning is achieved by the system organization". In that sense, it is similar to its target system, "(…) given that in real language children do not receive constant feedback about what is incorrect in their speech” (Li & Zhao, 2013, pp. 1–2).

The example reveals that models are far from being exhaustive depictions that have a complete correspondence with their target systems. Apart from being unpractical, this idea neglects that models require a selection of degrees and certain criteria for representing. Not all models represent, and the ones that represent do not do this in the same way. In that sense, models are not mere simplifications of more complex realities. While substituting, they are also changing what is at the scope. Models such as SOM are not just representations of phenomena, otherwise, they would consider all the relevant aspects purposely ignored in order to make these models work. Besides, if SOM models were representations, then they would target unsupervised language learning systems, which are hypothetical entities created in the process of constructing these models, and different from the processes by means of which humans acquire language.

Models are artificial and concrete entities designed to make certain features salient and others inconspicuous. A psychoanalytic model of depression, for instance, stresses the role of early childhood in the development of this mental illness (Negele, Kaufhold, Kallenbach and Leuzinger–Bohleber, 2015), while a cognitive–neurobiological approach emphasizes the relation of rumination in depressed subjects to task–negative dominance (Marchetti. Koster, Sonuga–Barke and De Raedt, 2012, p. 243). The emphases in early childhood and in the feeling of a lack of attunement with certain activities point out to different things, but this does not imply that one model is false and the other true, nor that one of them is more comprehensive than the other (regardless the comprehensiveness of models can be compared based on their formal complexity). These accounts simply are interested in different phenomena. In certain cases, targeted domains can be related. For instance, first–person experiences is connected with brain activations, but their relationships needs the design of a new model, e.g. a first–person neuroscience that establishes some hypotheses about how these domains are related (Northoff and Heinzel, 2006).

Semantic views point out that models, and not theories or principles, achieve representation. They consider that theories "are better thought of as families of models rather than as partially interpreted systems" (French & Ladyman, 1999, p. 105). Models are non–exhaustive indirect representations that mediate between theories and phenomena. They, instead, select certain properties of phenomena, creating hypothetical target systems in this process. Besides, semantic views are interested in finding the constituents or defining features of representational relations, instead of examining the means and material conditions that make models powerful representational tools.

Semantic views hold that models represent target systems by maintaining relations of similarity or isomorphism. In other words, what constitutes or defines the representational function of models, that is, their capacity to stand for a target system is a relation of similarity or isomorphism. In what follows, these ideas and the criticism of pragmatic views toward them will be introduced.

According to similarity–based accounts, “a model *M* represents a target system *T* if *M* is similar to *T*” (Frigg, 2006, p. 60; Suárez, 2003b, p. 227). In principle, similarity and resemblance are exchangeable terms. However, Suárez (2003b, p. 227) treats similarity as some sort of generalization of several resemblances (Suárez, 2003b, p. 227). A weak point in this definition is that it does not tell anything about which similarities matter when generalizing in order to establish a representational relation.

Ronald Giere, the main proponent of similarity, describes models as “(…) idealized structures that we use to represent the world, via resemblance relations between the model and real–world target systems” (1988, in Godfrey–Smith, 2006, p. 726). In his account, similarity is considered an important – perhaps defining – characteristic of scientific models. He nonetheless does not claim that models are similar to real–world things, but to target systems. The difference between real–world things and target systems is important in philosophy of science. Holding that models target real things involves defending a naïve realism, while the other option implies either a constructivist or an idealistic view. In the latter, scientific knowledge produces their objectivity, called phenomena.

In any case, similarity does not involve a full correspondence between the properties of models and real systems. Similarity is rather a characteristic of what is selected as being relevant into question; otherwise, models would be replicas of real systems. Similarity–based accounts hold that models are idealized systems that resemble real systems in certain respects and degrees (French & Ladyman, 1999, p. 110). Giere assumes a perspectival realist of models in which they stand for target systems that exist in the world, not artificial or constructed entities created in scientific research. As a result, abstract models must be empirically connected to real systems in the world (Hackenberg, 2009, p. 395; Giere, 2010, p. 271).

Processes of interpretation of models and identification are needed in this context for establishing a connection between models and real systems. Interpretation in this context involves connecting abstract principles of theories with physical properties of entities, and identification consists in associating the elements of models with those of target systems. According to Giere, it is the task of scientists, not of philosophers of science, to explain how these processes are conducted (2010, p. 271). Because of that, in his account perspectival realism is presupposed rather than being justified (French & Ladyman, 1999, p. 111).

Critics of similarity–based accounts argue that similarity is not a constituent of representation because it is asymmetrical relation, whereas representation is not (Giere, 2010, p. 274). If *A* is similar to *B*, then *B* is similar to *A.* But if *A* represents *B,* that does not imply that *B* represents *A.* In other words, similarity is a property shared by two or more entities which resemble each other, while in representational relations only sources are directed at targets, not vice versa. In Suárez’s words, “a source is not represented by a target merely in virtue of the fact that the source represents the target” (2003b, p. 232).

Giere intends to overcome the limitations of similarity by adopting an intentionality–based account. According to this, “agents intend to use models to represent a part of the world for some purpose” (2010, p. 274). Since similarity needs to be invoked or stated by some agents in certain contexts of inquiry, similarity is no longer considered an intrinsic property of the relation between models and target systems. Certainly, some sources appear as more useful for representing certain things in the world than others (e.g. artificial neural network models seem more useful than other models for representing biological connections). Giere (2010) does not deny this, he instead stresses the need of intentions and purposes for instantiating a relation of representation based on similarities. Intentions determine which similarities must be taken into account and the success of certain models when representing (2010, p. 275). This explanation solves the problem of asymmetry because the agent establishes the directionality of similarity between models and target systems.

Similarities are not strong structural correspondences between models and target systems. As a result, accounts based on them tolerate different representational vehicles. Models can be graphical, pictorial, combine mathematical with non–mathematical features, etc. Since there are several possibilities for establishing a relation of similarity, this concept is regarded as vague or ambiguous. Frigg (2006) remarks, “everything resembles everything else in any number of ways” (p. 61). In that sense, it is unclear how intentions provide good criteria for determining the relevant respects and degrees that make similarities representationally success, nor of the contexts in which models and target systems are similar (Frigg, 2006, p. 61).

Furthermore, similarity–based accounts ignore the fundamental roles of dissimilarities in representational relationships (Suárez, 2003b, p. 231). Models are not representationally successful only for being similar to their target systems, and certain dissimilarities are essential to them. Consider, for instance, the dissimilarities between a scale model and its target. If the size of a scale model is not dissimilar to the size of the target system, it would not be a scale model in the first place. Dissimilarities are implicit in the definition of certain models; thus, they are not what make representations unsuccessful; this would be the case if representations were mere replicas of targets.

To summarize, these accounts argue that scientific models represent target systems by maintaining relations of similarity with them. They also stress that agents establish which similarities matter. By assuming this, they apparently give an answer to the problem of asymmetry, although they are unable to explain in which degrees and respects similarities matter. These limitations should not lead to a complete denial of the idea of similarity since, in some sense, it appears as naturally or intuitively acceptable, but it needs a stronger justification than these accounts.

Isomorphism–based accounts, in contrast, do not provide intuitive explanations of how models represent. Besides, these accounts do not consider that agents' intentions are the only criteria for representing. These accounts regard isomorphism as a property of the "(…) relation between a model and its target system" (Frigg, 2006, p. 53), that can be defined as a mathematical correspondence between the elements and relations of models and those of target systems (Da Costa and French, 2000, p. 119). Knuuttila (2005) emphasizes that in these accounts isomorphism is closer to the general notion of isomorphism than to a narrow isomorphism (see also Bueno & Colyvan, 2011, p. 349). Considering this, this section follows Frigg in not examining the different isomorphism–based accounts (embedding, partial isomorphism, or homomorphism) (2006, p. 53).

In order to be isomorphic to a target system, a model must have a structure that represents another structure in the target system. According to Frigg, "the structure *S* represents the target system *T* if *T* is structurally isomorphic to *S* and *S* is intended by a user to represent *T*" (2006, p. 54). As in the case of similarity, isomorphism does not implicate a complete correspondence of models and systems structures, but only one of certain respects and degrees. In that sense, few parameters are selected in the isomorphic relation, not the complexity of target systems' mechanics (Suppe, 1989, p. 94).

Frigg remarks, "isomorphism assumes that the target exhibits a structure, but in the context of certain description" (2006, p. 55). This means that target systems' structures are not a real property of phenomena, but artificial entities that are created or invoked by agents in scientific domains. Despite agents invoke these structures and determine the parameters of isomorphism, they do not influence the matching between these structures. If the parameters vary, the isomorphic relation might change or even disappear. Additionally, choosing parameters can be challenging since – formally – isomorphism can be multiply instantiated. Suárez (2010) explains this idea in the following terms:

> (…) since there are always different ways of *cutting out* its domain of elements and relations, every physical object instantiates simultaneously several structures. The physical world underdetermines its mathematical structure – which may only be ascribed under a particular description (p. 96).

According to this, isomorphic structures need first to be instantiated. To do this, models must be divided into parts and relations between these elements which correspond to parts and relations in target systems (Suárez, 2010, p. 96; 2016, p. 452). Isomorphism–based accounts must justify why certain instantiations are more appropriate to describe certain target systems than others. These selections are not arbitrary. Scientists decompose models attempting to provide coherent descriptions of target systems (Frigg 2006, p. 59). Since scientists pursue exact or – at least – reliable knowledge, isomorphism could be regarded as a strategy for accomplishing their efforts.

Despite isomorphism is instantiated, rather than being an intrinsic property of the relation between models and target systems, proponents of isomorphism do not refer to it as one possible type of representation, but as the defining characteristic of scientific representation (Frigg, 2006, p. 59). Nevertheless, “models involve, but are not reducible to structures” (Frigg 2006, p. 53). In other words, isomorphism does not cover all the possibilities of models for representing phenomena, and models include several components not tied with structural relations. Even in the case of formal mathematical models, mappings from sources to target systems are not purely structural because there are “additional pragmatic and context–dependent features in the process of applying mathematics” (Bueno and Colyvan, 2011, p. 352). Although the isomorphism between mathematical models and their target systems enable the first to be used to make inferences about their target systems, these models can also be used for pragmatic purposes that are not related to isomorphism such as unifying various disparate phenomena (Colyvan 2001, 2002, as cited in Bueno & Colyvan

2011, p. 351). In that sense, isomorphism should be considered only as possible representational strategy, instead of reducing scientific representations to isomorphism.

Isomorphism is a symmetrical relationship, whereas representation is not (it is unidirectional). Besides, isomorphism can be independent of representations. Frigg (2006) observes, “neither one of a pair of isomorphic objects represents the other. Two copies of the same photograph, for instance, are isomorphic to one another but neither is a representation of the other” (p. 54). Certainly, scientists’ intentions direct how models represent target systems, determining which elements are isomorphic. However, this answer appears as “(...) a paraphrase of the problem rather than a solution” (Frigg, 2006, p. 54). Indeed, isomorphism becomes irrelevant if only depends on agents’ intentions of representing by using models. Things can be isomorphic to other things in a very general sense. In that regard, isomorphism would be at best a method to regulate representations “(...) by imposing constraints of what is the admissible” (2006, p. 55), without reducing representations to isomorphic structures.

A shared intuition in semantic vies is that models do not represent pre–theoretical worlds, but phenomena already interpreted in certain ways. They are not directed at things in the world, but to data models (Frigg, 2006, p. 59), and “(…) theory is not confronted with data but with models of data, constructed in sophisticated and creative process” (Van Fraassen, 1985, p. 271). Thus, a relation of isomorphism is not established between sources and external or independent systems, but between sources and empirical systems or models of data (French & Ladyman, 1999, p. 112). A theory, in this regard, is already a system with certain structure or a set–theoretical structure" (French & Ladyman, 1999, p. 116). As a result, models do not directly represent raw data. Rather, target systems are organized as data models, that is, either already formalized empirical sets or abstracted entities that already possess certain structures. Considering this, French and Ladyman (1999) conclude that "the use of isomorphism and related notions is perfectly legitimate" (p. 113).

In sum, isomorphism–based accounts state that models' structures represent target systems' structures. These structures are instantiated in processes of interpretation that rely on scientists' intentions that determine the direction of representation and which aspects of both entities are isomorphic. However, the pragmatic dimension of scientific representation cannot be reduced to a relation of isomorphism, which is unable to explain why certain models are better for representing than others, or why minimal models can be sometimes more useful than other models depending on the context of inquiry.

A last challenge for semantic views is the problem of misrepresentation, which has been formulated by Knuuttila (2010) as follows:

> (…) the isomorphism account does not accept false representations as representations. The idea that representation is either an accurate depiction of its object which is interpreted in terms of isomorphism within the structuralist conception – or it is not representation at all, does not fit our actual representational practices" (p. 143).

In other words, semantic views maintain that a model is either similar or isomorphic to its target system, or it is not. These views cannot handle the success of representations in terms of degrees of accuracy, utility, employability, etc., claiming instead that partial similarity or isomorphism are already similarity or isomorphism, respectively, whereas misrepresentation are not representations at all. Pragmatic accounts, in contrast, tolerate misrepresentation because they believe that representation is a triadic relation composed by a source, a target, and an agent. Knuuttila and Merz (2009) affirm that “the pragmatic approach to representation could be seen as a critique of the structuralist notion that is part and parcel of the semantic conception of models” (p. 148). According to pragmatic views, scientists assess the success of models by considering the means in virtue of which they achieve representation. Consider the case of misrepresentation presented by van Fraassen (2008) in the following:

> *Misrepresentation* is a species of representation after all: a caricature of Mrs. Thatcher may misrepresent her as draconian, but it certainly does represent her, and not her sister or her pet dragon or whatever else she may have. Yet even if we take the caricature to represent her because of some carefully introduced resemblance there, we can declare it a misrepresentation by insisting that it represents her *as something she is not*. A caricature may represent a rather tall man as short (as a well–known cartoon depicts Supreme Court Justice Clarence Thomas as very small compared to the chair he occupies), but it represents that man, and not someone that it resembles more as to height. A caricature misrepresents on purpose, to convey a message that is clear enough in context but is to be gleaned in a quite indirect fashion (p. 14).

The case of caricatures is illustrative of how misrepresentations function. In caricatures, certain characters are exaggerated in comparison to others. Despite their inaccuracy, they can represent. Thus, a caricature is not an accurate representation, but a representation that stresses and makes salient certain qualities. It can be inferred from this that accuracy is not a constituent of representational relationships;

otherwise, misrepresentation, reductionist representations, etc. would not be representative at all.

Pragmatic approaches go beyond the dyadic relation of correspondence between source–target, or model–target systems, by introducing agents into this relation (Knuuttila, 2010, p. 143; Suárez 2003a; Giere, 2004). Agents, nonetheless, are elements of these relations in a different way than sources and targets. They instantiate this relationship, making it possible in the first place, with their intentions to represent using models. Besides, they create the models and select the material vehicles and means to represent. Knuuttila (2010) describes the role of intention in the following terms:

> They create the directionality needed to establish a representative relationship: something is being used and/or interpreted a model for something else, which makes the representative relation triadic, involving human agency (p. 143).

Agent’s intentions are prerequisites for sources to be directed at targets. These are acts of stipulation which enable representation in the first place. In that sense, “(…) pragmatic approaches make representation less a feature of models and their target system than an accomplishment of its user” (Knuuttila, 2011, p. 265). In the case of scientific representation, users expect to gain knowledge of target systems by means of making inferences about the target systems by using models. Then, if the study of a model's behavior does not allow to reasonably infer anything in the target system, then the model fails as a representation, but it still is a representation.

Misrepresentation is also different from simplifications in models’ design. In facial coding, for instance, algorithms are trained to recognize emotion via clustering of facial expressions, gestures, body language, tone of voice, etc. By doing this, artificial agents can detect anger, contempt, disgust, fear, joy, sadness and surprise. Since “the overall usability of computational templates is based on their generality and the observed similarities between different phenomena” (Knuuttila, 2010, p. 146), cognitive scientists expect to gain knowledge of these phenomena through these models. This detection is, however, only partially isomorphic to the way humans recognize emotion since, for example, it does not take into account other factors such as language that facilitate the recognition of certain emotions – although these models could be trained to consider other information from different modalities, too. The task of recognizing emotions has been consciously impoverished for increasing predictive power by only considering seven basic emotions. This is a reductive strategy that facilitates the recognition of certain emotions but makes the model unable to represent more complex emotions such as envy, jealousy or anxiety. Hence, recognition of emotions in the model and in humans are different phenomena, and the only valid inferences that can be drawn from the model to humans are related with human capacity to infer others’ emotions based on their facial expressions, which is not the same as recognizing emotions.

A pragmatic approach to modeling is able to justify why reductive accounts such isomorphism or similarity are valid scientific postulates. The idea that an agent intends to use models to represent parts of the world for some purposes does not involve any sort of subjectivism. Pragmatic approaches claim that certain actions enable representation. In this context, Suárez distinguishes the constituents from the means that facilitate representation. Constituents define what is a scientific representation, whereas the means are context–dependent characteristics of sources that make them useful representations of their target systems (2010, p. 93). Accounts based on similarity or isomorphism do not explain the means of representation and are unable to explain from where the representational power of sources comes from. Besides, they define representation in terms of something else (reductionism), i.e., they take a characteristic of certain representations as the property that defines any representation, forgetting the means and concrete aspects of representations equally involved. In that sense, “(…) pragmatists doubt of the existence of a substantive philosophical analysis of scientific representation that could account, for, on a general level, how and in virtue of what models give us knowledge” (Knuuttila, 2011, p. 263).

Deflationary views are at odds with substantialist accounts like isomorphism or similarity. They assert that representations are dependent on certain contexts of inquiry, essentially linked to their use and, “(…) is thus best characterized by its function or role in the practice of model–building” (Suárez, 2010, p. 96). The representational power of a source depends on model construction. Modeling practices, and not merely acts of stipulation, are what makes models enough powerful for representing their target systems (Suárez, 2010, p. 98). Suárez resists a naturalistic view of representations since “(…) he resists saying anything substantive about the supposed basis on which the representational power of representative vehicles rests” (Knuuttila, 2005, p. 1264).

Although pragmatic approaches based their account on models' power to represent, they are too vague for explaining how models are epistemically successful since they neglect that representations are defined by any constituent. In other words,

> whereas the strong representationalism accounts fail to present an adequate notion of representation and impose too strict success criteria, deflationist accounts remain too minimalist to assess the epistemic value of model (Knuuttila, 2011, p. 264).

In order to assess how models provide knowledge about their target systems, pragmatic accounts consider the construction of models. In this context, they realize that "model construction happens *before* the possible real target systems (Knuuttila, 2010, p. 140). In other words, target systems' existence is not prior to their materializations or instantiations that initiate when scientists stipulate that a model targets certain system. Target systems can be unknown or merely hypothetical, and only through modeling practices they become observable phenomena. In contrast to semantic views, pragmatic views have non–realist implications since representation is not prior to the material vehicles and subjective intentions that instantiate it. However, if target systems are not prior to the representational relation, how are they created? Regarding this question, Knuuttila (2010) stresses that

> (…) instead of directly trying to represent some selected aspect of a given target system (…) modelers proceed in a roundabout way, seeking to build hypothetical systems in the light of their *anticipated results* or of certain general features of phenomena they are supposed to exhibit (p. 146).

In other words, modelers use their previous knowledge of the presumed target systems such as beliefs, conjectures, intuitions or even previous scientific backgrounds to create these hypothetical systems. Considering this, the question of how they gain knowledge through models cannot be answered by a mere appeal to isomorphism or similarity, and it is related to the expertise of modelers.

In order to explain how knowledge is gained in modeling practices, Suárez adds a special characteristic in modeling in science: "(…) the source must have the capacity to be employed by an informed and competent user to draw valid inferences regarding the target – what is known as 'surrogative' reasoning or inference" (2010, p. 98). This type of reasoning postulates a certain kind of internal structure inherent to sources. Source structures can be decomposed into parts, and these parts can be interpreted as the corresponding parts in target systems, while sources structures are the relations between the parts in target systems. This implicates that if a modification or certain behavior occurs in a source, it can be fairly expected to occur in a target system. Certainly, only informed and competent agents can draw valid inferences from target systems using models (Suárez, 2004, p. 773; Knuuttila, 2005, p. 1264).

Finally, Suárez differentiates the directionality of models, a property by which they denote their targets, from intentionality. Since he understands intentionality as a defining feature of mental or cognitive states (2010, p. 98), he treats it as a characteristic of an isolated agent that is opposed to collective practices. Nevertheless, collective practices are meaningless without agents that interpret them. In addition, the *aboutness* of mental states is not only constituted in isolation but precisely in intersubjective engagements. Suárez does not realize the complementarity between intentionality and intersubjective engagements because his account of intentionality is based on theories of mind and in Brentano's philosophy, which lead to solipsistic views such as "(…) scientific model represents via someone's mental state" (Suárez, 2010, p. 98). Moreover, he opposes intentionality, which is "in the mind", from collective practices which are "in the social world" (2010, p. 99). The next section will confront this solipsistic view by reviewing phenomenological accounts of the intentionality of theoretical acts.

Summing up, what distinguished pragmatic from semantic approaches is the emphasis on agents' stipulations and expertise, and in the means that models need to represent.

Pragmatic approaches avoid attributing anything substantial to representations, e.g. defining them in terms of something else (similarity of isomorphism). However, "(…) once we introduce users into the relationship of representation, its explanatory power starts to fall apart (…) nothing very substantial can be said about it in general" (Knuuttila, 2010, p. 145). Indeed, semantic accounts recognize that agents play an important role in representational relations, but they intend to define representations in terms of something else. In contrast, claiming that representations are user–dependent without explaining what users add in this equation seems to be an unfruitful strategy.

Knuuttila considers that the limitations of pragmatic approaches can be solved by looking at models as epistemic artifacts, that is, "collective objects of knowledge (…) [that] mediate between different people and various practices" (2005, p. 1266). This means that they can serve various purposes apart from representation, as she emphasizes in these terms:

> I suggest that we should approach models as epistemic artifacts, that is, as intentionally constructed things that are materialized in some medium and used in our epistemic endeavors in a multitude of ways (Knuuttila and Voutilainen, 2003)”. As parts and products of our scientific (and other) activities, models are endowed with intended uses, one of which is representation. This is in line with the aforementioned pragmatic approaches to representation (2005, p. 1266).

Knuuttila pays attention to the materiality and a concrete character of models that enable agents to use them for several purposes. One of these purposes could be representational. For instance, a model can be used to simulate the behavior of certain system. They can be used to make indirect representations and to accomplish strategies of surrogate thinking if the concrete scientific practices permit them. “Most of the information models give us is indirect, a result of inferences of various kinds” (Knuuttila, 2005, p. 1269). In any case, she does not deny the representational capacity of models but points out that models are not intrinsically representational. In other words, they do not have pre–established representative functions.

Finally, although Knuuttila realizes that “the most interesting properties of models are due to the way in which intentionality and materiality intersect in their diverse use” (Knuuttila, 2005, p. 1266), her research addresses the latter and disregards intentionality (just as pragmatic views do). To compensate this, section 2.2. will analyze the type of intentionality that underlies scientific representation.

2.2. Phenomenology of theoretical attitude and acts

Phenomenology provides a more substantial account of the intentionality of theoretical acts in comparison to analytic philosophy of science. Intentionality is the core concept of phenomenological tradition since its origins. In what follows, using several texts both from Husserl and some recent phenomenologists, the intentionality and attitude of theoretical acts will be described.

According to Mormann, Husserl's phenomenology is a foundationalist project that seeks to ground scientific knowledge in an analysis of its formal conditions of possibility. In that sense, it resembles the epistemology of scientific knowledge of the syntactic views that the semantic approaches reject (1991, p. 64). According to these views, the task of a philosophy of science is to describe the formal logic that underlies scientific activity. Mormann asserts that semantic views replace logic with mathematics (1991, p. 65) but persist in this type of foundation. However, regarding Husserl's concept of logic, Mormann seems to forget that this concept is much closer to Kantian transcendental logic than to formal logic in its traditional sense. A transcendental logic, according to Kant, refers to the categories that make possible the representation of phenomena; in other words, they do not refer just to non–contradictory entities, but entities that can exist (2007, p. 100). This logic includes categories such as 'relationship', 'unity', 'multiplicity', 'causality', etc. In a nutshell, it is a logic for determining under which conditions certain objects that maintain relations that are determined by these categories can exist.

According to Husserl, a phenomenology of scientific activity takes a step further. It does not only analyze the kinds of relation that scientific objectivities maintain between them, it also describes the acts through which these objectivities are meant, and the formal categories that are involved in such acts. Phenomenology then is a descriptive analysis of the ways in which agents are directed at objectivities. Despite being involved in scientific activities, these objects are not perceived when doing science since scientists' attention is focused on the objects themselves, not in the structure of their experiences. These structures become a research topic through reflection, which brings a second–order awareness (Shim, 2011, p. 203).

The analysis of intentionality is bi–directional: it can be either i) an analysis of the structure of subjective experiences (what it is like to do scientific activities), or ii) an analysis of what is meant in scientific activities (intentional object). Husserl distinguishes the intentional object from the real object, i.e., a spatiotemporal thing. The intentional object is the object as it is intended, while a real object is the intended object (Husserl, 2001b, p. 113). For example, "a chair as perceived" is an intentional object, and "the chair" is the real object. Regarding perception, Husserl's view is close to realism since he believes that intentional acts like this give access to transcendent realities, i.e., things that are not in the mind of a subject but in the external world. Nevertheless, this access is not transparent by itself and it needs to be disclosed through a reflective analysis which he refers to as transcendental philosophy. He describes the idea of a transcendental philosophy as follows:

> (…) a transcendental philosophy in our definition (…) is a philosophy which, in opposition to prescientific and scientific objectivism, goes back to knowing subjectivity as the primal locus of all objective formations of sense (…) (1970, p. 99).

According to this, the strategy of phenomenology is to describe a subjective realm to disclose the objective formations of sense (meanings). Regarding scientific domains, scientific representations are objective formations or meanings, whereas scientific acts are the subjective realm. In *Ideas 2*, Husserl develops his account of theoretical acts and theoretical objects. What is meant in these acts, in the most general sense, are the concepts of nature and experience, which are described as follows:

> (…) nature, one would say first of all, is the total spatiotemporal "universe", the total domain of possible experience: thus (…) take the expressions "natural science" and "experiential science" as synonyms (Husserl, 1989, p. 3).

Any natural entity is spatiotemporal and an object that can be experienced; in other words, it is an empirical object. Theoretical acts are directed at nature, but Husserl distinguishes natural objects from all possible types of objects when referring to them as spatiotemporal realities (1989, p. 3). An imagined object – for instance, a minotaur or a gold mountain – are not part of nature in this sense because they are not spatiotemporal things. The category 'objectivity' is different from the category 'spatiotemporal' since not all predicates that can be ascribed to the first can be ascribed to the latter.

The idea of nature refers to what underlies natural sciences. In other words, it is what is meant in by sciences even in implicit form. Phenomenology describes how this idea is constituted in the experiencing subject: a consciousness which experiences natural science and thinks natural–scientific objects (Husserl, 1989, pp. 3–4). In other words, Husserl describes how scientific acts apprehend and validate their knowledge of empirical objects. Since this approach is phenomenological, it discloses the experiencing attitude that "(…) determines in advance what is or is not a natural–scientific object" (Husserl, 1989, p. 4). This attitude is theoretical in contrast to axiological and practical attitudes, whose predicates are not objects of interest in natural science.

The theoretical attitude belongs to a subjectivity that intuits and thinks in a natural–scientific way. By describing this attitude "(…) we will learn that what is termed 'nature' is precisely the intentional correlate of experience as carried out in this attitude" (1989, p. 4). Since nature is the correlate of this attitude, "nature is there for the theoretical subject" (Husserl, 1989, p. 4). This is different from affirming that nature is a concrete, evident or fully specified concept. Rather, nature is an ideal object in the sense that is "(…) an object of possible knowledge" (Husserl, 1989, p. 4). In other words, one can say that any physical object is part of nature, but when intending to represent nature without referring to any concrete entity, this concept simply cannot be grasped. In that regard, it seems more convenient to consider it a horizon or space of possible objects which excludes values, works of art, etc., rather than a concrete or material object.

Although axiological or practical objectivities are not given in theoretical attitudes, this does not imply that they cannot become objects of theoretical acts. These acts treat their intentional objects as spatiotemporal in the world when representing, judging or thinking them. In contrast, virtues – that is, the objects of practical attitudes – such as justice or moderation are not objects of theoretical interest by themselves. They are originally given in practical contexts concerned not in gaining an understanding of something but in acting in certain ways. But even such acts presuppose certain objects or, using Husserl's term, they are 'doxic' for being about something. An act of justice is directed at certain situation and depends on certain intuited value. Its difference, though, with a theoretical act is that its interest or attitude is not directed at the objectivity as such. Thus, doxic experiences are lived experiences oriented towards non–theoretical objects, but they in principle can become into theoretical acts.

In theoretical attitudes “such lived experiences are performed or carried out in the function of knowledge” (Husserl, 1989, p. 5). These experiences are not directed at the experiencing subjectivity, the one who perceives, represents, remembers, etc., but rather to what is perceived, remembered or represented. Husserl claims that these acts are performed in a certain function of knowledge, i.e., an active living experience in which there is an explicit apprehension of something. By explicit it is not meant that the object directly appears through these acts, but that the subjectivity explicitly appeals to its object while apprehending or judging it. Certain thing is represented, judged, etc. For example, when judging that the sun is bigger than how it appears to my perception is a theoretical act that apprehends the object ‘sun’ as smaller than how it appears to my senses. This is a theoretical object that has been previously constituted in the stream of consciousness since the sun is already an object before becomes the object of a theoretical attitude.

In theoretical acts “(…) what is objective becomes a theoretical object, an object, that is, of an actively performed positing of being in which the Ego lives and grasps what is objective, seizes and posits it as a being” (Husserl, 1989, p. 13). In other words, theoretical acts postulate the existence of their objects and characterize them as themes. This is a distinctive way than relating with objects in a natural attitude (Husserl, 1983, p. 57). This attitude is not reflective since it only cares if the object is spatiotemporal, without distinguishing the senses by means of which is given. In a natural attitude, for instance, phantasies are considered unreal because they cannot be perceived. However, objects of phantasy are not absurd for not being spatiotemporal.

Since natural sciences operate within a natural attitude, understanding what kind of knowledge is gained in that attitude could give insights regarding the nature of scientific representation. Phenomenologists often analyze the perceptual or intuitive accesses to the things in the world, i.e., those which directly present things in an original, non–derivative way. In Husserl's words:

> (…) presentation in general, by this, we always understand those experiences precisely making the objectivity they refer to presentational for authentic acts, positions taken toward something (2008, p. 276).

Does science deal with objects of intuition? Phenomenologists consider that scientists rather deal with derived or non–original objects. Specifically, scientific representations and models are far from being intuitive or perceptual accesses to their target systems, nor to present them directly. Besides, the source of the representational relationship is a hypothetical theoretical construct rather than a thing of perception since modelers do not perceive it, and rather they postulate hypothetical target systems. In sum, since the relation between the source and the target does not involve any intuitive access, scientific modeling operates with mediated representations. If this is the case, why this is a legitimate knowledge?

Acts whose objects are directly given in intuition are called by Husserl non–derived or original. Acts that, on the contrary, do not deal with directly given objects are called derived. They are grounded in non–derived acts and can occur as a modification of them. An object of the imagination such as a unicorn or a gold mountain is derived and can be represented despite not being an object of perception. A representation of an object of imagination is possible since its elements can be perceived individually. Gold and mountain are perceptual objects that can be stored in memory. Imagination can reconfigure and combine them, leading not to representations of perceptual objects, but precisely to imagined objects. A possible reply to this account could assert that these objects are also perceptual because someone could see them, for instance, in a children's cartoon, or even on television. However, imaginary objects are not limited to the internal states of a mind that imagines them, and can in principle be perfectly understandable when presented in an external medium. But without perceptual sources, objects of imagination are not possible. In this sense, phenomenology recognizes a genetic dependency of the latter to the first. Objects of imagination, then, are representations (*vergägenwertigung*). This is also the case of phantasy which "(…) is not itself a pre–sentifying but a re–presentifying representation" (Husserl, 2019, p. 317).

The last idea brings Husserlian phenomenology closer to fictional accounts to models developed in recent philosophy of science, which does not defend a realist view of scientific representation, but rather sees models as epistemic artifacts by means of which knowledge of certain processes is gained (Knuuttila, 2017). Nevertheless, theoretical attitudes do not refer to their objects as imagined entities; rather, they postulate their existences. Science has a commitment to truth and to the correspondence between propositions and states of affairs. However, these aspirations can be held without supposing that scientific knowledge has direct or non–mediated access to real things.

Derived acts have a specific type of validation or fulfillment. The content of a derived act, e.g., a proposition, is meaningful only in certain contexts. It needs to be embedded with a meaning intention that makes an expression significant (Husserl, 2001a, p. 193). A meaningful expression is different to what it refers to, that is, its sense and its objective correlate. Mere utterances do not make an expression meaningful because they need the intention to mean something. Their objective correlate is the meaning of what is expressed, which can be intuitively accessible or not since "(…) an actually given objective correlate, which fulfills the meaning–intention, is not essential to an expression" (Husserl, 2001a, p. 199). Thus, only the meaning of certain expressions can be fulfilled, but all expressions have semantics. Otherwise, symbolic thinking might be "insolubly enigmatic" (Husserl, 2001a, p. 209).

But if meaning does not come from intuition, where does it come from? Husserl (2001a) answers this question as follows:

> Expressions and their meaning–intentions do not take their measure, in context of thought and knowledge, from mere intuition – I mean phenomena of external or internal sensibility – but from the varying intellectual forms through which intuited objects first become intelligibly determined, mutually related objects. Thus, even if expressions can have a different non–theoretical function, symbolic intentions point to categorically *formed* unities (p. 199).

In order to be meaningful, propositions need to be structured in certain ways. Husserl denominates ‘intellectual forms’ the categories that structure sentences. These categories underlie the semantics of empty expressions and make them understandable. For instance, the statement “a computer is on the table” contains the categorical forms ‘a’, ‘is’, ‘on’ and ‘the’, which are needed for this sentence to be meaningful. These forms not intuitively perceived but combined with terms that designate thing in the world can be used to mean something. They do not refer to mere objects but to situations or states of affairs. The adequate combination between categorical forms and designators turns expressions significant. Inadequate combinations such as “the table on is computer a” are meaningless.

In sum, meaningful propositions are embedded with meaning intentions and possess categorical forms that articulate them. However, they are empty intentions by themselves. Husserl (2008) distinguishes them from fulfilled ones as follows:

> One can also set up (*aufstellen*) a contrast splitting the same distinction differently, namely, contrast (*gegenüberstellen*) the intuitive, as it were, full presentations (*Vorstellungen*) and, on the other hand, the empty presentations to which all merely significative, mere verbal presentations, for example, belong. In the one case, they are appearances of their objects, in the other, not (…). Empty intentions are directed toward objects, but are not “authentically” presentations of them. They do not make an object “stand before us”, just appear. Authentic presentations set the object before or portray it (*stellt den Gegenstand vor oder dar*). Namely, they objectively apprehend material given in sensation and illusion (…). Hence, here people speak of apprehending, or even of representation (*Repräsentation*). Unauthentic presentations do not do that. Their objects are not represented, not portrayed (*dargestellt*) in content conscious in the sensation or imagination. This content is not considered as an object, not indicated as an object (p. 276).

The idea of fulfilling a sentence is related to the embodiment of an expression. Any expression, in order to become meaningful, must be accessible to the senses or at least represented in certain way. Distinguishing empty intentions from intuitive presentations do not imply that the first cannot be fulfilled. Scientific representations are also empty intentions that postulate the existences of their objects, which need to be verified in order to fulfill these intentions. As stated, it seems that all intentional attitudes tend to fulfillment but this is not necessarily the case. If empty intentions are directed at non–testable objects, they will not be fulfilled. Besides, if an empty intention is directed at an imagined object that resembles in certain respects to a target system, then the fulfillment is not proven by demonstrating the existence of the imagined object, but by demonstrating the resemblance between them. Husserl describes the fulfillment of objects whose existence has been postulated in these terms:

> With respect to apprehending, fulfillment is an identification, however, not only that, but also a *verification of the position–taking*. The unauthentically presented objective moments occur in the fulfillment process in the quality of givenness, and this quality confirms the original belief in accordance with the intentional moments overshooting the original givenness. In advancing from perception to perception within an essentially coherent context of perceptions, the belief acts or belief expectations are confirmed over and over, and with this an ever more far–reaching consciousness of givenness of the object is constituted, i.e., a consciousness of its being as an ever more fully realized consciousness of being (2008, p. 308).

It can be inferred from this description that fulfillments are not immediate apprehensions of what is meant and rather are temporal and involve a constant confirmation of beliefs' expectations. In the case of scientific representations, this confirmation also involves intersubjective validation, and most beliefs are just taken for granted rather than being constantly validated. “When we operate with concepts that do not refer us to any kind of experience we have had, we are following along with stable tabs or reference that freely circulate in the field of culture” (Chernavin, 2016, pp. 57–58). In other words, representations in most cases belong to a shared world of meanings, which is described in the following terms:

> World–representation, thing–representation means here: what is represented as such in my and our human representing. It is not until I have taken up the ultimate transcendental standpoint and have grasped from it the infinity of transcendental all–subjectivity – that which finds itself in the world and finds itself as living into the world in worldly, subjective experiences – in its totality, that this tension vanishes, and the difference between representation and actuality vanishes (Husserl, 2019, p. 600).

Without going into details of Husserlian transcendental phenomenology, the emphasis on the transcendental suggests that meanings are not constituted in isolation and depend on an intersubjective horizon in which beliefs are taken for granted or considered commonsense, while others are constantly updated or even emerge. This means that in principle beliefs do not need to be individually confirmed at each time. This is also true for scientific knowledge, which in principle can be tested over and over. Accepted representations and beliefs become parts of a common horizon of meanings, and that is why they are taken for granted rather than tested at every step of scientific procedures, despite their truth or validity are not intuitively given.

Unlike representations that are embedded in shared cultural meanings, the theoretical attitude of scientists postulates the existence of states of affairs. The act of postulating the existence of something has a judicative nature because it asserts that objects have certain qualities rather than others. Any hypothesis is grounded in a judicative act that is an element of any scientific procedure. This act is described as follows:

> More precisely stated: Judging is meaning and, as a rule, merely supposing that such and such exists and has such and such determinations; the judgment (what is judged) is then a merely supposed affair or complex of affairs: an affair, or state–of–affairs, as what is meant (Husserl 1982, p. 10).

Judgments are directed at states of affairs and they need to be validated, otherwise, they remain mere empty intentions to mean something. In the same way that some beliefs' expectations are proven and others that remain untested, or presupposed, there are also immediate judgments and mediated judgments that presuppose the firsts (Husserl 1982, p. 10). Scientific knowledge is supported on a tradition that in no case starts from scratch, which is not based only on immediate evidence but operates with judgments that, in turn, depend on other judgments, etc. Science, as a tradition, does not depend on a single subject that verifies its judgments, but on an intersubjective horizon. In other words,

> (…) every theoretical formation then takes on, in the intersubjective nexus of the human community, as such a part of intersubjective science, a manifold of worldly relations yet: a relation to the first scientific discoverer and its real documentation for the objective tradition, to the different pupils and their original acquisition, and so on. (Husserl, 2019, p. 353).

From a Kuhnian perspective, it can be argued that scientific tradition does not possess a linear continuity and rather scientific concepts are incommensurable or have no absolute meaning, that is, one independent from a set of theoretical beliefs of a scientific community (Suárez, 2003a, p. 266). This idea implies that intersubjective validations are possible only in the context of a certain tradition. However, Husserl also points out that scientific knowledge can be traced back into certain principles whose evidence can be reactivated. Thus, these principles are preserved in traditions but their evidences are independent from them. Since they are commonsense rather than directly intuited, “science is a mediate cognition, inferential or deductive” (Husserl, 2019, p. 413). Hardy (1992) explains this idea as follows:

> Scientific knowledge, then, according to Husserl, is grounded knowledge. It is achieved through a demonstration in which the fact to be scientifically known is “deduced” from antecedent conditions. For any state of affairs that we would claim to know scientifically, we must be in a position to show how it necessarily follows from other states of affairs. If, in turn, we are to claim to know these antecedent states of affairs scientifically, we must also be in a position to show how they follow from other states of affairs. The process of grounding knowledge can be repeated. But it must eventually terminate in certain ’principles’ if it is to avoid either an infinite regress or circularity (p. 7).

In this way, scientific knowledge is composed of both immediate and mediate judgments, and the relationship of dependence between them can be elucidated. Mediated judgments are evident if the judgments that support them become evident. Principles, therefore, are judgments whose evidence does not rely on other judgments. Considering this,

> Such principles must be “immediately known” if they are to function as the first principles of science wherein all other propositions of the science are grounded. They themselves are “groundless,” if by “groundless” we simply mean that their truth is not apprehended on the basis of other propositions. Thus, the very idea of science contains within itself the distinction between mediate and immediate judgments. Mediate judgments are ultimately grounded in a deductive fashion in immediate judgments. The immediate judgments are not grounded in other judgments, but in the direct intuitive experience of the states of affairs corresponding to them (Hardy 1992, p. 7).

At this point, it seems necessary to highlight that mediated knowledge is not less reliable or truthful. Truth, understood as correspondence between a proposition and a target reality, is a quality that belongs to mediated judgments and not of the presentation of things. In the context of empirical sciences, phenomenology realizes that there is no perfect adequacy or fully evidence of scientific representations; rather, they are inductive, fallible and only approximate. In Hardy's words:

> Due to the contingent character of the laws of empirical science, the open–ended character of the inductive process by which they are confirmed, and the ineluctable margin of error in all observations, the laws of the empirical sciences are not only tentative, but approximate. The kind of knowledge they afford, when compared to knowledge in the strict sense defined by the classical idea of science, is knowledge only a in "wider, modified sense" (1992, p. 29).

This limitation, however, does not situate scientific knowledge at the same level as common knowledge or opinion. On the contrary, empirical sciences are approximate regarding an ideal of perfect knowledge. In this way, there is no sharp distinction between scientific and non–scientific knowledge despite they can be distinguished considering how they are acquired. Phenomenology bets for a continuity of them in which science is considered a privileged cognition that has means of foundation that cannot be found in knowledge based on experience. In this sense, phenomenology – unlike enactivism – can provide an account of the continuity of forms of knowledge, which comprises both abstract knowledge and the most basic forms of sensible intuition thanks to the distinction between mediated and non–mediated forms of knowledge. In Husserl's words:

> We must also take note of the fact that what we call theoretical or scientific knowing is only a privileged higher form [of knowledge] that relates back

> to lower levels – for example, to the various forms of sensuous intuiting and sensuous imagination, with the sensuously intuitive modes of judgment belonging to them, which not only historically precede scientific judgments as typical forms of the cognitive life of pre–scientific humanity (and indeed are already to be found in animals) but which also play a role in scientific thought itself as an always and necessarily co–functioning basis and underlay (2019, p. 48).

The different forms of knowledge do not exclude each other. This suggests that elements of pre–scientific domains can be involved in scientific activities. Similarly, the meanings given in the scientific world have an impact in the non–scientific domain.

This section has presented the accounts of scientific representation of two philosophical traditions: analytic philosophy of science and phenomenology. In conclusion, the characteristics of the scientific representation that can be expected in representations of cognitive sciences are outlined in what follows:

- Models, not theories, are the means of scientific representation.
- Scientific representation is composed of a source and a target system.
- Similarity and isomorphism are scientific strategies for explaining how models stand for their target systems.
- The relevant respects and degrees of similarity need to be specified.
- The parameters of isomorphism need to be explicitly stated.
- Representations are accomplishments of their users.
- Representational power comes from the model construction in which targets systems are unknown hypothetical systems.
- Model construction is prior to target systems.
- Scientific representation operates with surrogative reasoning, i.e., what occurs in the model is expected to occur in the target system.

- Empirical sciences are directed at nature, i.e., the set of spatiotemporal things.
- Theoretical attitude has a function of knowledge that explicitly intends to apprehend its object.
- Theoretical acts postulate or posit the existence of their objects and characterize them as themes.
- The theoretical object is not an object of intuition or perception.
- Theoretical acts are derived and not original acts. They are composed of meaning–intentions and intended objects. They can be fulfilled or not.
- Scientific activities do not require a constant validation of their principles; they are rather presupposed.
- Scientific representation is inductive, fallible and approximate.

3. Representations in Computational Views of Cognition

In a research article, Lewandowski et al. (2019) ask why the opinion of a minority of denials of climate change has had a strong effect on the public opinion despite the scientific consensus about the negative impact of CO_2 emissions in climate change (around 97% among domain experts). Although there is evidence of well–organized campaigns of elitists groups to influence public opinion on climate change (e.g. analysis of IRS data estimates the income of a network of conservative think tank near $1 billion annually) (p. 125), Lewandowski and his colleagues are not directly interested in such empirical data, but in modelling these social interactions through computational simulation that represent them using idealized models. By doing this, they expect to explain the dynamics of particular social interactions. This task is done through the construction of an agent–based model of three groups of actors: scientific community, operatives of the organized denial network, and the public. "All actors are represented by rational Bayesian agents that seek information by inspecting climate data or by communicating with each other" (p. 125). They find that "(a) unbiased agents necessarily acquire belief in the climate–change hypothesis even from an initial position of skepticism; (b) to persist with denial agents must be biased; (c) the presence of such biased agents can delay, but not prevent, belief formation in the scientific community; (d) the presence of contrarian voices, especially when disproportionally represented, can prevent the public from acquiring the scientific consensus position” (p. 135).

Instead of directly approaching social interactions, simulating brings new possibilities of inquiry. Certainly, modelers acknowledge the limitation of agent–models whose behaviors are highly idealized (they do not fully match nor represent the behavior of real agents). Nevertheless, this strategy makes possible a quantification of the influence of certain agents to others’ beliefs by exploiting the characteristics of Bayesian networks. In this context, they simulate how rational agents progressively accept the scientific evidence of climate change, and how overrepresented information prevents the public from agreeing with scientific consensus. Cognitive scientists often employ these and other computational models and simulations for studying cognitive phenomena. The kind of knowledge gained through them is the topic of this section, which discusses under which conditions computational models can represent cognitive processes.

Apart from using methods such as experimentation or brain imaging, cognitive scientists are nowadays highly dependent on computational modeling. Computational models are mathematical tools aimed to study or simulate the behavior of a variety of systems. This is a mainstream definition that does not distinguish computational models from computational templates, the broader cross–disciplinary syntactic structures used in different scientific domains (Humphreys, 2002, 2004). Neural networks, as it will be discussed later, should be considered templates rather than models, and require the addition of several constraints in order to represent phenomena.

In section 2, scientific representation was described as a relation involving sources, target systems and agents. If cognitive scientist use computational models to understand cognitive processes, then it needs to be specified whether computation i) is equivalent to cognition (target system), or ii) it is just a part of the scientific framework by which scientist make sense of the mind (source). This section bets for the second alternative by analyzing how scientists use models for simulating and representing cognitive phenomena.

This section is divided as follows: the first part introduces the computational–representational understanding of the mind (CRUM) and the computational theory of mind (CTM), which are two different versions of the well–established idea in cognitive science that cognition is, or at least involve some sort of, computation. These views underlie two main approaches in cognitive science: cognitivism and connectionism, respectively. This section introduces the ways these approaches understand representation. After this, it analyzes how connectionist models can achieve representation by considering the processes of analogical reasoning, idealization, de–idealization, etc., occurring in the construction of models. It is argued that by themselves neural networks are highly idealized templates, similar to black–box units, unable to represent phenomena. However, if certain conditions are met, they can be used to achieve representation.

3.1. Computational Views of Cognition

Cognitive scientists frequently assume that cognition is, or at least involves, some type of information processing. The assumption behind this is that “the essential levels of the cognitive system’s organization are best described as information processing” (Milkowski, 2013, p. 26). How exactly information processing occurs is still a matter in dispute in this discipline, whose answers divide proponents of cognitivism, connectionism, and the 4Es approach. Despite the divergences of these approaches, they agree in considering cognition as something different from mere mechanical reactions to stimuli (Bermúdez, 2014, p. 8). To give an example, cognitive psychologists treat human cognition in terms of mental processes in which sensory inputs are transformed, reduced, elaborated, stored, recovered and used (Neisser, 1967, p. 5, as cited in Casey and Moran, 1989, p. 144). Information flows over the stream of human cognitive processes, it is labeled, classified and sequenced to support actions.

Examples of cognitive processes include, but are not restricted to, communication, decision making, perception (in a broad sense), social skills and tool use. These processes are often studied by various disciplines due to their broad character. Scientists and philosophers of science have realized that interdisciplinary collaboration is essential to gain understanding of these complex phenomena

(Gallagher and Zahavi, 2012, p. 1). Following this thought, cognitive scientists use findings and methods from anthropology, artificial intelligence, biology, linguistics, neurosciences, philosophy, and psychology (Thagard, 2005, ix), and other disciplines because in certain sense almost any discipline is related to the mind (Bermúdez, 2014, p. 6).

Although cognitive processes are significantly different among each other, cognitive scientists use similar models, templates, approaches and methods to study them. For example, artificial neural networks are used for modeling natural language processing, image recognition, and so forth. Although modelers approach them using models with similar structures, they are non–correlated dissimilar phenomena. If the same organizational principles and techniques are used for modeling different types of phenomena, it can be questioned whether the supposed structures behind these phenomena exist as independent objects from the representational relation in which they occur. This idea will be discussed in detail in the next section.

The advent of digital computers has been perhaps the most influential event for the study of the mind. In its origins, cognitive science treated the human mind as an information processing machine similar to computer programs that process and manipulate information (Casey & Moran, 1989, p. 144). Indeed, a central assumption of cognitivism – the initial and most influential paradigm of cognitive science – treats the mind as an information processing machine and the brain as a computing machine which manipulates symbols that *stand for* or represent objects and/or states of affairs in the world (Bickhard and Terveen, 1995, p. 11). More precisely, cognition is regarded as similar to how digital computers processes data by following rules that map “input strings of digits, plus possibly internal states, to output strings of digits” (Piccinini and Scarantino, 2011, p. 3). Despite big success of these approaches in many respects, they fail in providing an adequate account of some fundamental questions such as: How does meaning arise? How can a machine represent significant events?

As it has been stated, cognitive scientists assume the cognition is or involves processing of information. In order to explain how processing of information works, they describe it either in terms of a computational–representational understanding of the mind (CRUM), or a computational theory of mind. Thagard refers to CRUM as the central hypothesis of cognitive science, which can be summarized in these terms: “thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures” (2005, p. 10). CRUM is sufficiently broad for covering both symbolic and non–symbolic forms of representation as well as local and distributed representations (Zhang and Patel, 2006).

CRUM states that minds possess mental representations analog to data structures and that computational processes are similar to algorithms (2005, p. 11). These analogies allow scientists to describe cognitive processes as computational processes, using computational models to simulate cognitive phenomena. CRUM, however, is not committed to the metaphysical hypothesis that cognition is computation. Not even Thagard supports this hypothesis, but only that this analogy (computational metaphor) is fruitful for the study of the mind. By endorsing the analogical view, he avoids attributing the characteristics of certain models to the phenomena they stand for. This is not the attitude of other philosophers of mind and cognitive scientists that assume that minds compute (Bermúdez, 2014, p. 13).

The success of the analogy between minds and computers is undeniable. Several scientific disciplines have emerged influenced by this analogy. In cognitive psychology, which is one of them, “the advent of digital computers offered (...) both a plausible metaphor (i.e., the mind as a computational system) and a new method (i.e., computer simulation) for the investigation of the mind” (Casey & Moran, 1989, p. 144).

Computational views have been the mainstream approaches to cognition since the foundations of cognitive science. Nowadays, the situation appears to be slightly different but only at first glance. Although certain proponents of the 4Es approach propose non–computational and/or anti–representational approaches to cognition, i.e. modeling cognitive phenomena using dynamical systems, these alternatives are far from being fully–accepted systematic research programs.

Although it is assumed that cognitive science is an interdisciplinary enterprise, cognitive psychology, a discipline referred as “the marriage between psychology and artificial intelligence” (Núñez et al., 2019, p. 7) is overrepresented in comparison to other disciplines such as anthropology or biology (Núñez et al., 2019, p. 4), i.e., there are significant differences concerning the proportion of papers of cognitive psychology (Gentner, 2010, as cited in Farkaš, 2012, p. 423), as well as the amount of computational–oriented research and their founding in comparison to other approaches (Chemero, 2009, p. 16). As a result, it can be asserted that the current situation is not entirely different from the origins of cognitive sciences regarding the primacy of computational–oriented approaches.

The overrepresentation of computationally–oriented research is far from being neutral, influencing scientists' assumptions regarding the nature of cognitive processes and the role of information processing. To give an example, some proponents of constructivism hold that there is an analogy between artificial intelligence and human intelligence on the basis that no substantial differences between them can be demonstrated (Nehaniv, 1999, p. 2). They assume that despite the differences in the implementation of these intelligent system, their underlying information processing structures are the same. In consequence, they expect that they gap between these types of intelligence can be reduced by creating artificial intelligent systems inspired by living organisms, while holding the metaphysical assumption that the mind is a computational machine (Samuels, 2018, p. 106).

CRUM does not need to be committed to this metaphysical assumption since its engagement with the computational metaphor is only due to its pragmatic advantages for understanding the mind. Farkaš (2012) refers to computational modeling as an indispensable tool in cognitive science, asserting that "all physical systems whose variables can be measured, can be viewed as computational" (p. 405). Computational models are useful as far as they permit the manipulation of measurable physical systems; more precisely, at least certain aspects of cognitive processes are measurable, and computational models can inform about them. Without reducing cognition to computation, what is argued is that the most prominent way to approach the measurable dimensions of cognitive phenomena is by using models that exploit numerical and symbolic data.

The computational theory of mind (CTM), in contrast, endorses the metaphysical view of Cognition as computation, defining the latter in terms of “a computational process defined over linguistically structured representations” (Sprevak and Colombo, 2018, p. 1). The differences between CRUM and CTM are very subtle and are often ignored (these terms are sometimes interchanged). Nevertheless, they can be distinguished because CRUM treats computation only as a useful method of representing cognitive phenomena for pragmatic reasons, whereas CTM holds the metaphysical assumption that cognitive systems are computational machines.

According to CTM, cognitive processes are computations. In virtue of this, they can be best described through symbolic representations and manipulations of syntactic rules of digital computing processes (Garrido, 2010, p. 41). CTM endorses the cognitivist or symbolic model of the mind, in which “the mind is a symbol system and cognition is symbol manipulation” (Harnad, 1990, p. 336). Kelley (2003) describes how exactly minds manipulate symbols in the following:

> The computer can take a series of symbols as input. These symbols are representations of some other concept or construct (which actually have meaning only to the human operator). The computer can then manipulate these symbols by using some pre–set instruction set. It can then output a result of the symbols based on the previous manipulation process. So, if a computer is given the number ‘4’ and instructed to add the number ‘4’ to the number ‘7’, it will output the symbolic result ‘11’ (pp. 848–849).

Kelley emphasizes that CTM operates with pre–established rules and symbols. This makes it more restricted or less general than CRUM because it does not include sub–symbolic and non–ruled based computations used by connectionists models. Additionally, CTM asserts that mental processing is computation is an empirical hypothesis rather than a metaphor or a fruitful analogy for studying the mind (Pylyshyn, 1985, p. 55). This supposed empiricism is still supported by some recent proponents who consider as empirically validated the view of the nervous system as a computing machine that can be described in mechanistic terms (Milkowski, 2018, p. 516). This empiricist thesis is explained as follows:

> Computational processes in the physical word have two salient properties: (1) their structures can be described in terms of a computation, and (2) they are physical. Hence, the intuitive idea that many philosophers and physicists endorse: a physical process is computational if, and only if, there is a computation such that the physical states the process consists of correspond to states of the computation in a one–to–one fashion. After all, a true computational description of a process must correspond to reality. This correspondence is usually framed in terms of isomorphism: the structures involved have to stand in identity relation to each other (Milkowski, 2013, p. 29).

Following this, an isomorphism is established between a computational description (source) and a physical computational process (target system). Although this hypothesis seems to be empirically testable, physical computational processes are phenomena that have been already interpreted in terms of computational systems. Thus, when the hypothesis is proven, what is tested is only that a computational model fits with the interpretation of a physical system in computational terms. Since non–computational interpretations of the same physical system are also possible, what is demonstrated is only the correspondence between a computational description and a computationally interpreted phenomenon. Regardless of the accuracy when testing hypotheses, the empiricist thesis cannot be proven. It can be stated, instead, that isomorphism in this context is just a construction assumption, instead of a property of the target system.

As stated, CTM is connected to cognitivist or symbolic approaches to cognition. Harnad (1990), the symbolic model of the mind states that "symbol strings (…) captures what mental phenomena such as thoughts and beliefs are" (p. 336), whereas "(…) the symbolic level (for them, the mental level) is a natural functional level of its own, with ruleful regularities that are independent of their respective physical realizations" (p. 336). From the 1960s to the 1990s, CTM and cognitivism played a major role in cognitive science because it offered an empirically testable and practical–oriented approach to the mind. In the words of Varela, Thompson and Rosch in *The Embodied Mind* (first published in 1991):

> Cognitivism has the virtue of being a well–defined research program, complete with prestigious institutions, journals, applied technology, and international commercial concerns. We refer to it as the center or core of cognitive science because it dominates research to such an extent that it is often simply taken to be cognitive science itself. In the past few years, however, several alternative approaches to cognition have appeared. These approaches diverge from cognitivism along two basic lines of dissent: (1) a critique of symbol processing as the appropriate vehicle for representations, and (2) a critique of the adequacy of the notion of representation as the Archimedes point for cognitive science (2016, p. 8).

Nowadays, cognitivism is no longer the mainstream paradigm in cognitive science. There are at least two other paradigms that attempt to be a unifying theory of cognition apart from it: connectionism, and embodied dynamicism (Thompson, 2007, p. 4), also referred as the 4Es approach. While connectionism puts into question the primacy of linguistically structured representations and syntactic rules (Sprevak & Colombo, 2018, p. 1), enactive criticism is deeper in the sense that it rejects the computational metaphor by arguing that it is unable to explain the gap between computational states and consciousness (Thompson, 2007, p. 3).

The idea that cognition is computation over symbolic representations implies that symbols stand for or represent things in the world. Steels (2008) points out a widespread confusion regarding the term ‘symbol’ in cognitive science. The notion of symbol in symbolic programming languages “is a pointer to a list structure containing a string known as the ‘print name’” (p. 228). Thus, it lacks of any sort of semantic content. The confusion is that they are not distinguished from meaning–oriented symbols in the philosophical debates (p. 228).

However, since cognitivists views claim that human minds compute, at least certain properties between computer systems must be manifested in human minds. The difference is that human minds compute with semantic contents. To be successful, cognitive acts rely on the accuracy of their representations (Varela, Thompson & Rosch, 2016, p. 40). Cognitivists think that representations involve semantic contents and syntactic structures. Semantic contents are what makes representations meaningful, while syntactic structures articulate and organize the interactions of meaningful contents. Only syntactic structures are therefore shared by computers and humans). Computations, in cognitivism,

> (…) are operations on symbols that respect or are constrained by those semantic values. In other words, a computation is fundamentally semantic or representational–we cannot make sense of the idea of computation (as opposed to some random or arbitrary operation on symbols) without

> adverting to the semantic relations among the symbolic expressions (Varela, Thompson & Rosch, 2016, p. 41).

As stated, symbols are mere strings of characters and need to be interpreted by an external human agent in order to signify something. Thus, if cognition is reduced to computation, then it is unrelated to meanings – at least in computers. Since CTM does not only pretend to describe the cognition occurring in computers but in explaining mental representations in the physical world as well (Milkowski 2013, p. 138), it is expected that at some point the symbols used by computers are similar to the symbols used by human computations (Newell, 1980, p. 136, as cited in Milkowski, 2013, p. 139). However, there has not been any successful account of this possibility.

Milkowski distinguishes two notions of a symbol in computer science:

1. “A symbol is a token, i.e., a piece of information that a computer processes.
2. A symbol is a pointer to a list structure” (Steels, 2008, p. 228, as cited in 2013, pp. 139–140).

According to the first notion, symbols are formal and lack content and referent. A meaning can be arbitrarily ascribed to a token by an external observer. Thus, a computational system cannot operate with representations – in the sense of mental contents – by itself. In the second notion, a list structure contains relevant data such as “symbol’s name, temporarily assigned value, a definition of function associated with this symbol” (2013, p. 140). These pointers can be used to access information indicated by some other symbols. In both cases symbols lack meaning, there are no external referent they target, and meaning is at most an attribution of an external human observer. A difference between computer and human minds is that symbols in the brain or mental representations “do not seem to require any *external* interpretation to be meaningful (2013, p. 144). Milkowski concludes from this that CTM “cannot really account for the emergence, or constitution, of representation” (2013, p. 144). Enactivists agree with this criticism emphasizing the limitations that syntax must handle semantics, describing how cognitivism explains this as follows:

> A digital computer, however, operates only on the physical form of the symbols it computes; it has no access to their semantic value. Its operations are nonetheless semantically constrained because every semantic distinction relevant to its program has been encoded in the *syntax* of its symbolic language by the programmers. In a computer, that is, syntax mirrors or is parallel to the (ascribed) semantics. The cognitivist claim, then, that this parallelism shows us how intelligence and intentionality (semantics) are physically and mechanically possible (Varela, Thompson & Rosch, 2016, p. 41).

Following these authors, cognitivism intends to reduce semantics to syntactic structures analog to cognitive processes. Thus, cognitivism overemphasizes the role that these structures play in cognition, seeking the computational programs that formalize certain cognitive acts. Indeed, it privileges the role of functional properties (software) and dismisses non–formal properties (hardware, embodiment). In principle, the idea is that algorithms can be instantiated or embodied in many different media (Steels, 2008, p. 235). In that sense, cognitivism depicts cognition as a computing process that is hardly related to meanings and in which embodiment is unessential.

Ignoring the importance of the physical instantiations of cognitive processes is a commonplace in CTM, according to which “human cognition takes place independently of the physical or physiological characteristics of the system” (Jorna, 1990, p. 275). While CTM emphasizes that cognition does not dependent on their physical instantiations, cognitivist approaches are only interested in the functional aspects of human cognition that can be described as software or computer programs (Jorna, 1990, p. 275). Connectionism, in contrast, does not provide a disembodied view of cognitive processes, as it will be discussed in the following.

Cognitivist and connectionist models differ in the way information processing operates in them. Cognitivist models treat cognition as functions operated over symbolic representations. In order to grasp “the fluidity and adaptability of human information processing” (McClelland, Rumelhart & Hinton, 1987, p. 3), one must find the right computer program or software without considering the particular embodiment. Computer programs depict possible cognitive processes. Given that cognitivism situates cognition at the functional level, tasks such as deciding, planning, playing a game or classifying elements are considered cognitive processes regardless of whether an artificial or human agent does them.

A weak point of cognitivism is that human information processing and computer models information processing are not similar. Even accepting the functional argument, the material differences between human and computational cognition are too conspicuous for not taking them into consideration when building models that represent human cognition. Apparently, this is not the case of connectionist models which are inspired by the “computations” occurring in the brain. However, they are often very abstract and, in many cases, idealized models of real cognitive processes. Bermúdez explains the idea of connectivity in these terms:

> The brain is an extraordinarily complicated set of interlocking and interconnected circuits. The most fundamental feature of the brain is its *connectivity* and the crucial question in understanding the brain is how distributed patterns of activation across populations of neurons can give rise to perception, memory, sensory–motor control, and high–level cognition (2014, p. 211).

According to connectionism, lower–level interactions can be distinguished from higher emergent orders. Higher–level phenomena such as decision–making, perception, tool–use, etc., have their roots in lower–level interaction without being reduced to them; in other words, they have distinctive traits and their own laws of behavior (Humphreys, 2009). Lower–level cognitive domains are, for instance, the patterns of activations by which the neurons are interconnected, or the embodied engagements of cognitive agents with their environments. Farkaš (2012) asserts that "considerable empirical evidence, covered by the umbrella of grounded (…) Cognition, suggest that higher Cognition is embodied in the lower–level sensorimotor process" (p. 414). Despite higher cognitive domains emerge from these interactions, this does not imply that the firsts can be explained by accounts that solely refer to the lower level (for example, by reducing cognitive processes to processes occurring in the brain).

Artificial neural networks (ANN models) are capable of simulating lower–level interactions such as real connections occurring in the brain. Indeed, they are regarded as psychologically plausible for being inspired by these interactions (McClelland, Rumelhart & Hinton, 1987, p. 11). These models are taken into account due to their physiological flavor, "since they seem so much more closely tied to the physiology of the brain than are other kinds of information–processing models" (McClelland, Rumelhart & Hinton, 1987, as cited in Rusanen and Ylikoski, 2007). Therefore, these accounts propose a similarity between ANN information processing and computations in the brain as far as both systems operate with “interactions of multiple simple processing elements or units that send excitatory and/or inhibitory signals to other units (McClelland, Rumelhart & Hinton, 1987, p. 10).

Furthermore, the architecture of an ANN resembles aspects of biological functioning and neural activations in the brain. Nevertheless, “most artificial neural networks are not biologically plausible in anything but the most general sense” (Bermúdez, 2014, p. 72). ANN models can serve to represent cognitive processes only if certain specific conditions are achieved, and they need to in order to accomplish the general aim of connectionism of disclosing the abstract principles governing brain’s functioning (Bermúdez, 2014, p. 72).

According to Clark and Lutz (1992), ANN consists of

> (…) a large number of simple processing elements (nodes) connected together to form a network. Each connection in such a network has an associated weight (or strength) which determines how important that connection is and how much influence the nodes connected by it can have on each other. The values computed by the nodes, and which are passed between them via the connections, are all purely numerical. (…) For a given architecture (i.e. which nodes are connected together and for a given choice of computation performed by the nodes) this really amounts to choosing a set of interconnection weights which will make this possible (p. 1).

This general characterization of ANN models is sufficient for the purposes of this section. According to this, ANNs are composed by nodes, weights and activation functions. Nodes are sets of units that can be positive or negatively activated. These units are separated into layers whose activations are determined by the activations of all their single units. There are three different types of layers: input, hidden and output layers (in feedforward ANN models). Input layers are made up of input units "which receive inputs from sources outside the network". Hidden layers contain weighted inputs, and output layers are made up of output units, which send signals outside the network" (Bermúdez, 2014, p. 73). The activation of the different units in the layers determine how information is processed in a network. This is a parallel processing since its outputs are determined by independent multiple units (Bermúdez, 2014, p. 72). As Clark and Lutz point out,

> The process of training a network is somewhat lengthy. It is usual to begin with a random assignation of weights and then present the network with a training series of input patterns of activation, each of which is associated with a target output pattern of activation. The input patterns are presented. Differences between the actual output pattern and the target output pattern result in changes to the weights. This is what the learning algorithm does – adjust the weights in order to reduce the difference between actual and desired output (1992, p. 74).

ANN models are not systems that follow pre–established rules. However, as in the description of Clark and Lutz, they can be trained via supervised learning,the other two possibilities are unsupervised and reinforcement learning. All these forms of learning are biologically relevant and important. For instance, when explaining the computations occurring in the cerebellum, Doya (1999) argues that supervised learning modules in the cerebellum are used as internal models of the environment, whereas unsupervised learning models provide representations of internal states and environmental states.

ANN models learn how to map certain inputs to certain outputs. They can do this by modifying the weights until desired outputs are produced. Far from being explicitly ruled, learning consist in the adjustment of connections' strengths (McClelland, Rumelhart & Hinton, 1987, p. 32). The capacity to learn is another similarity between ANN models and human minds. Regarding this, McClelland, Rumelhart and Hinton (1987) proposed an analogy between ANN models and human minds considering that humans are smarter than computers because their computational architectures simultaneously consider various pieces of information at the same time (p. 3). This means that their architectures operate with parallel distributed processing (PDP), which "offers alternatives to serial models of the microstructures of cognition" (p. 12).

Instead of storing information using symbolic representations, ANN models store information in microfeatures. They are the "atomic elements in a distributed connectionist representation" (Sharkey, 1991, p. 146), and stored in hidden units which do not contain anything semantically interpretable (i.e., fragments that can be putted together to compose a symbol), and depending on the activations they can be differently combined in order to construct something that can be meaningful for an external observer. For instance, when processing images ANN models decompose them into patterns that are sets of derived–pixel–based features which are later used to reconstruct the imagens (Egmont–Petersen, de Ridder and Handels, 2002, p. 2280).

Computational models can use either local or distributed representations. Local representations stand for individual items using single units (i.e. any symbol), and distributed representations use patterns of activation through various units to stand for things. While local representations are transparent in the sense that each unit is clearly labelled, distributed representations are opaque because their units are only understandable in the patterns of activation (Sharkey 1991, pp. 144–145). In ANN models, representations are mostly "distributed across the microfeatures represented by several different hidden units at the same time" (Dawson, 2005, p. 174). Since microfeatures are not localized symbols, "knowledge (…) is not stored in the connections of a special unit reserved for that pattern, but is distributed over the connection among a large number of processing units" (McClelland, Rumelhart & Hinton, 1987, p. 33). Symbolic information emerges from the large collection of these unlabeled units (Sharkey 1991, p. 147).

To summarize, CRUM and CTM are different versions of the general attempt of using computational models to represent cognitive phenomena. They differ from each other because the first is a pragmatic approach to the use of computational models, while the second is committed with a metaphysical view of cognition as a form of computation. These two approaches underlie cognitivism and connectionism. CTM endorses a symbolic–based approach to representation, while CRUM can be regarded as a pragmatically oriented approach to computer models.

3.2. How do computer models represent?

As it has been stated, cognitivist models are symbolic and connectionists models mostly sub–symbolic as far as they use microfeatures that do not stand for things by themselves (Dawson, 2005, p. 174). Due to this characteristic, sub–symbolic models are more similar to nervous systems' function than symbolic systems. However, ANN models are very abstract and only under certain circumstances they can be useful for representing. This section discusses some representational strategies involved in modeling using ANN.

According to Morrison and Morgan (1999), the process of constructing models is essential for making them useful for representing. Whereas there are no settled rules for model construction as in other scientific practices (p. 12), and they "are partially independent from both theory and data (or phenomena)" (p. 14), models can function as instruments for representing phenomena (p. 25). Morrison and Morgan point out that models do not have an inherent representative function, but they can enable representation if a model has a relation of dependency with a theory or a phenomenon (pp. 25–28). Although ANN models are inspired by real connections occurring in the brain, this is not enough for such relation. They differ from for being highly idealized entities with fewer connections than biological ones. Besides, they can have negative activations, which is not possible for biological connections. These differences do not make them useless for representing these phenomena. In any case, modelers need to add certain constraints when building these models to enable them for a representative function. Consequently, if ANN models are constructed for representing cognitive phenomena, modelers must take into consideration

> (…) the constraints of psychological phenomena, and neuroanatomical and neurophysiological phenomena systems. (…) Using neural constraints can be a good modeling strategy even if these constraints are not correct in their details (Stinson, 2018, p. 120).

It is important to keep in mind that model construction is a process. Models do not need to have all these constraints from the beginning, nor their representational capacities. Neural networks are abstract, highly idealized entities, and by adding constraining in the process of constructing them, they become more similar to empirical systems. Without the addition of constraints, neural networks remain as mere computational templates (Humphreys, 2002, 2004).

The addition of constraints can be regarded as a representational strategy that can be applied to computational templates in order to make them useful for mediating between theories and phenomena. These constraints, in turn, depend on the phenomena of interest. If, for instance, a model intends to represent certain activation occurring in the brain, negative activations (a possibility of ANN models) should be excluded as far as this characteristic is not typical of them. The addition of constraints can also support an inferential use of models.

In the process of constructing models, there are several negotiations concerning their abstraction or concreteness, and their dissimilarities or similarities with phenomena. Models do not always need to be similar to phenomena, but this is one of their pragmatic possibilities. Adding physiological constraints, for instance, can "increase the strength of the inference" (Stinson, 2018, p. 124). In this case, a model is more similar to a target system and this makes possible to use it to infer properties on this target. In sum, constraints are used to determine and specify the ANN model in response to the phenomena of interests. This is a different sense of how ANN models are inspired by real cognitive phenomena. For instance, ReLU function (i.e. y =

max(0,x)) is inspired by real neural activations that do not have a negative value. This function suppresses any activation that does not fit with the way real connections behave. However, it ignores the fact that biological fact that real neurons saturate (they cannot be active above certain threshold), which is considered by other functions such as the logistic sigmoid.

While the similarities between ANNs and real connections can be maintained, the process of constructing these models can go a step further by de–idealizing the ANNs by considering the current knowledge of the phenomena of interest. Therefore, the addition of constraints can be regarded as a representational strategy to make a model more similar to real connections.

Scientific representation, as section 2.1 stated, cannot be reduced to a relation of similarity. But the addition of constraints can be rather related to an inferential use of models. In other words, it should serve for informing about an unknown domain and not just for representing the current knowledge scientists possess. This characteristic is called the inferential capacity of models, and Suárez (2004) describes it in these terms: “A represents B only if (i) the representational force of A points towards B, and (ii) A allows competent and informed agents to draw specific inferences regarding B” (p. 773). Considering this, if an ANN model is constructed considering the available knowledge of the target system, correct inferences about the latter could be drawn.

The addition of constrains do not make an ANN model more “similar” to its target system. In the case of ReLU function, the suppression of negative activations does not imply that an ANN model is more similar to what attempts to represent because the two systems are still very different from each other. ANN models are radically different in a material sense to the cognitive phenomena they represent or simulate.

As mentioned, constraints such as anatomical or physiological details added in the process of building models do not make models similar in a physical sense but can make them useful for representing phenomena. According to Stinson (2018), the constraints added to target systems reduce their abstraction, capturing certain mechanisms that can underlie them. In that sense, they can be useful for making inferences via using models (p. 127). Finally, the skillfulness and competency of modelers are manifested on how they make explicit the assumptions when constructing models. Otherwise, they would inflate the extent of the inferential capacity of their models.

Mechanisms in target systems, for instance, are postulated in ANN models when they are constructed following biological and/or physiological constraints (Stinson, 2018, p. 124). Following this assumption, ANN models and their target systems share the same kind of information processing and perform computations (Stinson, 2018, p. 126). However, this statement is too general to capture the conditions for representing. The addition of constraints supports the discovery of mechanisms since

> (…) the point of implementing networks roughly analogous to neural structures is to discover and explore the generic mechanisms at work in the brain, not to deduce the precise activities of specific structures (Stinson, 2018, p. 121).

According to this, the constraints of ANN models inspired in real phenomena are useful for representing mechanisms in these phenomena, and not just for implementing structures that serve other purposes (for instance, gaining predictive power). The emphasis on mechanisms can be regarded as a realist assumption of connectionist accounts. Mechanisms are not treated just as mere theoretical entities postulated following pragmatic drives but as the essences of target systems. As will be discussed, strategies of idealization involved in models' construction can employ minimal models to represent the basic conditions under which systems can operate. As such, minimal models that represent underlying mechanisms are more idealized than cognitive phenomena. When modeling the brain, for instance, "the point is evidently not to model the brain in detail, but rather to model the basic processing principles used by the brain" (Stinson, 2018, p. 127).

In comparison to formal algorithms, mechanisms are specific schemata of certain processes. As stated by Stinson, "a mechanism specifies both the algorithm plus the entities or parts involved in these activities, and their organization" (2018, p. 127). In that sense, a model that serves for representing also takes into consideration the different parts that compose a system. As in the case of the representation of the mechanisms, the representation of the parts of a system does not necessarily have to share physical characteristics with a target system and it could just reproduce their functions. In that sense, ANN models do not "assume localizability of functions" (Stinson, 2018, p. 128). They do not need to share resemblances of target systems' embodiment for sharing similar mechanism or parts. The notion of mechanism only points out that neural networks need certain structures in order to implement algorithms.

The realist commitments of connectionist approaches do not seem to be sufficiently justified for claiming that mechanisms underlie phenomena. Instead, they can be postulated as theoretical entities, similar to "(…) causes tending to produce like effects" (Stinson, 2018, p. 129). However, proposing the existence of mechanisms is not an arbitrary process. On the contrary, it occurs "when a regularity is discovered, such that one set of facts (…) tends to be followed by another set of facts" (Stinson, 2018, p. 129). Mechanisms are therefore discovered in the process of model construction by empirical observations and "only within certain ranges of parameter values" (Stinson, 2018, p. 129). Once mechanisms are postulated, it is expected that both models and target systems can instantiate them. Models could serve to explore the characteristics of the mechanisms in a well–known domain and under certain control conditions. After considering this, scientists use models inferentially to inform about the target systems.

Additionally, "one way of testing whether a property is an arbitrary or crucial feature of a mechanism's design is to see what happens if you remove or break it" (Stinson, 2018, p. 127). Taking into consideration the constraints from pathological or non–usual cases could serve not only to understand these phenomena but also to formulate hypotheses about the elements of normal cognitive systems and to understand their functions (Stinson, 2018, p. 127). ANN models can achieve this since they are resistant–to–noise flexible models. In other words, if their calibrations or their components vary, they can still work in more or less accurate ways. This is not the case of symbolic models that stop working if one of their elements change. Flexibility is what allows for accomplishing this strategy. For instance, "injuries can be simulated by modifying the network as a whole by adding noise, changing connection weights, or adjusting the learning rule" (Stinson, 2018, p. 128).

Furthermore, the inferential use of models is exploited in mechanistic models. This is useful when direct experimentations are not possible. For instance, if neurological subjects with specific injuries are not available (Stinson, 2018, p. 128), models can replace them. Other reasons supporting the inferential use are related to tractability. Since real systems tend to be very complex, they need to be simplified in order to make them tractable. In Stinson's words,

> Examining the human brain directly is invasive; experimenting on model species such as slugs, mice, or macaque monkeys face limitations and it cannot be assumed that these species process information in the way humans do. Computational modelling can be a strategy to overcome these limitations since they are able to model complex systems (2018, p. 121).

Computational models should not be regarded as mere reductive depictions of phenomena. In any case, reductive strategies are legitimate ways to gain knowledge of cognitive phenomena. The aim of classic AI was "to understand intelligence by precisely constructing a machine that reproduces the phenomenon" (Stinson, 2018, p. 121). But this attempt is in utopic if reproduction is understood in the strong sense of being committed with the metaphysical view of cognition as computation. The situation is different if it is rather understood as a pragmatically–oriented representation. In this regard, a practical possibility for models is to represent systems mechanistically, and to describe their components and mechanisms, parts, and interactions.

Furthermore, ANN models generally use bottom–up approaches that "(…) aspire to model the brain from the bottom up, starting by modeling brain anatomy and/or physiology in detail, like true simulations" (Stinson, 2018, p. 123). A bottom–up approach seeks first to determine the parts of a system and then to identify their functions. For instance, the brain can be regarded as a network composed of regions and connections, which can be treated as abstract nodes and connections in graphs (Stinson, 2018, p. 128). Compared to real neural architectures, ANN models appear as highly idealized systems. This has been asserted as a possible criticism towards these models. However, idealization can also be considered as a modeling strategy. Weisberg defines it as the "intentional introduction of distortion into scientific theories" (2007, p. 639). Distortion is introduced for making complex phenomena tractable and finding the minimal mechanisms that give rise to phenomena (2007, pp. 641–642; Stinson, 2018, p. 128). Thus, idealization is a reductive strategy that avoids unnecessary details in the attempt of explaining phenomena. As such, it can produce models that are too abstract for being representative. Nevertheless, models can also be de–idealized by considering the current knowledge of the phenomena of interest in order to give more accurate depictions.

Additionally, computational approaches employ analogical reasoning strategies. In sciences, they serve for gaining knowledge of an unfamiliar domains (i.e. cognition) via studying a more familiar one (i.e. computational models and/or templates) (Bailer–Jones, 2009, pp. 48–49). These domains are presumed to be similar and, in virtue of this, it is expected that efficient causes in known domains could give insights regarding causality in the unknown domain (Bailer–Jones, 2009, p. 55).

In cognitive science, computational and other models are familiar domains whereas cognitive processes are not, so the computational metaphor is a useful analogy to better understand the latter. Certainly, the analogy between computation and cognition is not always recognized as such. As it has been discussed, certain proponents of computational views rather claim that cognition is computation and, consequently, there is no metaphorical or analogical reasoning but a relation of isomorphism.

The inspiration of ANN models in real connections in the brain is not enough for isomorphism. As stated, the substantial differences in their particular embodiment cannot be grasped in isomorphism–based accounts, making them unrealistic. This is not the case of the analogical view which does not intend to defend a full correspondence between models and target systems. Instead, it treats ANN models as resources that guide scientists in formulating hypotheses and seeking the mechanisms that underlie phenomena (Bailer–Jones, 2009, p. 48). According to Bailer–Jones,

> A model could be an analogue, but this is not entirely the issue because the way to evaluate a model is not to judge whether it is analogous to something, but whether it, as it stands (analogous or not), *provides access* to a phenomenon in that it interprets the available empirical data about the phenomenon in a certain way (2009, p. 56).

As it has been stated, analogies do not involve structural correspondences and they are rather strategies to make sense of unknown phenomena. The knowledge of the familiar domain is used to organize and examine the unknown domain. The analogy between computation and cognition can serve, for instance, to predict certain behaviors in target systems, to postulate that they have certain properties in common and to divide them into parts. The following case exemplifies this:

> [children] are also often compared to computer program simulating the reasoning process. The assumption underlying such approaches is that we can learn about how humans reason ("how the brain works") from the way reasoning processes can be implemented on a computer. Again, an analogy is exploited in this very project: that between the real and the simulated process. In other words, the computational implementation is used as an analogue to the mental processes involved in analogical reasoning. This analogy is thought necessary because it is impossible to "inspect" the mind directly (Bailer–Jones, 2009, pp. 71–72).

This case reveals how analogies serve to gain understanding of unknown phenomena. In this analogy computer programs are the better–known controlled domain that is used to understand the less known domain of children reasoning. This analogy does not postulate any isomorphism, nor presumes that both domains operate with the same mechanism. The notion of mechanism, as it was stated, integrates connectionist processing of information with the material conditions for its implementation. If a model is excessively ideal or abstract, it would disregard these material conditions. This is a possibility for any ANN models since they use general empty structures (templates) and any further de–idealization relies on their construction. According to Knuuttila and Loettgers (2012),

> One characteristic feature of scientific modeling is the way modelers recycle equations, algorithms, and other formalisms around different domains and in which process the formalisms obtain different interpretations, depending on the domains they are applied to (p. 3).

These authors acknowledge that scientists from various disciplines nowadays use these empty structures for modeling purposes, which are directing the current interdisciplinary exchange of knowledge. As Knuuttila, Honkela and Rusanek (2007) point out, they "cross the boundaries of different scientific disciplines (…) through borrowing an already successful formalism from one scientific domain and applying it to another scientific domain". Humphreys refers to these structures as computational templates, which are syntactic computational structures that are used by scientists for different purposes (2002, p. 5), such as organizing scientific research. According to Lappi (2007),

> They are abstract computational schemata, the common syntactic core of a diversity of models, used to compute very different things in models of different phenomena. As purely syntactic abstractions, they can be considered in separation from any particular interpretation. Templates are not models of phenomena, but are instead necessary but not sufficient constituents of computational models in any given domain (p. 1226).

Humphreys considers that neural networks are cases of computational templates rather than models due to their degrees of abstraction (2004, p. 67). In comparison to models, computational templates are more abstract, idealized and closely tied to mathematical models (2002, pp. 2–5). Knuuttila and Loettgers (2012) add that they are "genuinely cross–disciplinary computational devices, such as functions, sets of equations, and computational methods, which can be applied to different problems in various domains" (p. 3). Other examples of computational templates include differential equations, statistical models, or computational models such as cellular automata and spin–glass models. Characteristic features of the phenomena they are directed at or a theory that underlies them do not determine all these templates. In that sense, they do not have a relation of dependency, which is needed for representing (Morgan & Morrison, 1999).

Considering this, ANNs must be de–idealized if representation is intended. The process of de–idealization can be conducted by considering the biological constraints of real networks. However, a relation of dependency does not lead to a unidirectional process. As stated, models are useful when direct experimentation is not possible.

Computational templates partially constitute computational models because they direct how they compute. Due to their abstraction, they require specifications of free parameters before being applied to models (Humphreys, 2002, p. 2). By themselves, as stated, they cannot be used for representing any phenomena, but their syntactic formulations enable phenomena to be computationally tractable (Humphreys, 2002, p. 3).

Whereas syntax is relevant for computational templates, they are also endowed with interpretations and justifications (Knuuttila & Loettgers, 2012, p. 4). In other words, despite they can be regarded as black–box units, "methodologically, the more important use of the templates comes when their construction is taken into account" (Humphreys, 2002, p. 5). Templates organize the processing of information in models, whose construction involves processes of approximation, idealization, and abstraction that make them computationally tractable.

Regarding idealization, Weisberg treats it not as a property of the relationship between a theory and a real–world phenomenon, but as the act of distorting theories or models, which is driven by certain representational ideals (2007a, pp. 639–640). He distinguishes three different types of idealization that respond to different ideals:

Galilean, minimalist, and multiple–models idealizations, which will be introduced in the following.

“Galilean idealization is the practice of introducing distortions into theories with the goal of simplifying theories in order to make them computationally tractable” (2007a, p. 640). It recognizes the complexity of targeted phenomena and creates simplified models that make problems tractable. Despite this can be seen as a reductive strategy, models’ simplification is a non–permanent practice. Scientists expect to de–idealize these models by “removing distortion and adding back details to her theories” (2007a, pp. 641–642), eliminating the reductive assumptions. The ideal pursued by this idealization is completeness, i.e., that models include all the relevant properties of target phenomena (2007a, p. 649), as in the case of mechanisms.

“Minimalist idealization is the practice of constructing and studying theoretical models that include only the core causal factors which give rise to a phenomenon” (2007a, p. 642). Minimal models are constructed considering only the factors that are crucial for the emergence of a particular phenomenon (2007a, p. 642). As a result, these models are extremely simple and do not consider the features that make models more realistic. What these models represent is rather “the interactions and structures that really make a difference, or the core causal factors giving rise to the target phenomenon” (2007a, pp. 642–643). Causal factors are the ones whose removal “prevents the model from entailing the phenomenon’s occurrence” (2007a, p. 643).

In the construction of minimalist models, any non–different–making factors (or not causal) should be omitted for the sake of simplicity. Because of this, minimal models can accept false statements that simplified their target systems. This idealization is also related to the idea of mechanisms in the sense that it captures “some single property or set of properties” (2007a, p. 644), capturing patterns and isolating causal factors (2007a, p. 645). Minimal models pursue the ideal of simplicity, which can be summarized as including the fewer elements as possible while maintaining a qualitative match between a model and target system (2007a, p. 650).

Finally, multiple–models idealization “is the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon” (2007a, p. 645). Scientists use multiple models when a single model is unable to provide high fidelity predictions, accuracy, precision, simplicity of phenomena due to their high complexity (2007a, pp. 646–647). Further, while using a minimal model for disclosing causality, scientists can additionally use several models for bringing more concrete details to certain phenomena, attempting to maximize predictive power without being necessarily interested in causality (2007a, p. 648). Multiple model idealizations are driven by the ideal of maximizing precision and accuracy, even if science is also interested in explaining why systems behave the way they do, and not just in providing black–box models (2007a, pp. 652–653).

As stated, idealizations should be complemented with de–idealization (i.e. using correction sets refine these processes considering available data) (Humphreys, 2002, p. 6), that make them more accurate depictions of phenomena. In any case, ideal models and templates need to be specified when constructing models. The addition of constraints can be regarded as a strategy of de–idealization, but other de–idealizations or specifications are also involved, as Knuuttila and Loettgers (2012) point out in the following:

> The construction assumptions of the computational template consist of an ontology, idealizations, abstractions, constraints, and approximations. An ontology specifies the kinds of objects referred to by a model. The correction set, in turn, is linked to the construction assumptions in that it relaxes some of the idealizations, abstractions, constraints, and approximations made and thus determines which parts of the model are intended to be interpreted realistically. Complemented with all these components, a computational template converts into a fully–fledged model (p. 4).

Knuuttila and Loettgers (2012) assert that de–idealizations occur in the processes of constructing models, which produces the representative function under certain circumstances. If models remain highly idealized, they are unable to represent. ANN models can be more or less abstract, and they can remain non–representative if they are driven and designed with intentions non–related with representing. Without de–idealizations, ANNs remain as mere computational templates that, in principle, do not serve for representing any cognitive phenomena. This argument appears as counterintuitive because it is taken for granted that the inspiration of ANN models in real connections (i.e. in the brain) is a point in favor of considering them as better depiction of certain cognitive processes, or even that

> the perceived similarities between different phenomena are really produced by the same kinds of mechanisms. We take it that a great deal of scientific modeling aims to study the mechanisms that produce natural and social phenomena (Knuuttila & Loettgers, 2012, p. 5).

As it was mentioned, the notion of mechanism presupposes that models represent by virtue of sharing structural similarities with their target systems. This property is expected to make them better depictions of cognitive phenomena (in comparison to symbolic models). These structural similarities are reflected not only in the constraints added into these models, but also in certain presumed features of the target systems such as distributed representations.

Considering this, it appears as reasonable to choose ANN models for depicting real connections occurring in a nervous system disregarding other models. But this is due not to their similarities to these systems, but because of what they offer for representing these phenomena. Similarly, scientists' strategies rather than an isomorphism between ANN models and cognitive processes are what make these models useful for making inferences about target systems. Since models have other possibilities rather than representing, which can either employ inaccurate data or present false statements, the conditions for representing must be specified. When a model has a representative function and points to a phenomenon, its dynamics, details, etc., modelers assumptions must specify the extent of the representational relation.

Knuuttila and Loettgers (2012) point out that computational templates such as ANN networks are interpreted, specified and corrected in the process of constructing models by considering available knowledge about target systems. As discussed, the notion of mechanisms is tied with the representative function of models that attempt "(…) to depict the basic mechanisms underlying some specific phenomena in a certain domain". Computational templates are not neutral because they are also inspired by certain phenomena. Since ANN models are inspired by real connections in the brain, they might be fit better with these similar phenomena than with other phenomena that could be better depicted with another representational vehicle.

Considering this, Knuuttila and Loettgers (2012) identify a "tension inherent in modeling practice between capturing the components and interactions of the supposed real causal mechanisms operative in the world and using general templates as a means" (p. 16). Following this, computational templates might only serve for modeling specific types of phenomena; for instance, those who can be explained by certain mathematical models. In that sense, the representative function of ANN models has a limited extent. Analogical reasoning could disclose the possible domains to which apply these models since scientists need to choose which templates are appropriate for representing certain phenomena and which are not. This idea is further developed as follows:

> Analogies and metaphors can rather be treated as devices that also contribute to the justification of a model in allowing the introduction of successful computational templates and modeling methods to a new field. Through these methods and techniques, the model gets some initial built–in justification (Boumans, 1999) (Knuuttila & Loettgers, 2012, p. 18).

Summing up, computational models are not representational by themselves. While traditional views of the use of computational models defend this idea by appealing to some variety of isomorphism, recent approaches to ANN models suggest various other possibilities of computational modeling not tied with representation. However, they do not deny that these models can be used for representing. According to these views, computational models can represent phenomena only if certain conditions are met. In order to enable the for representing, an analogy between computation and cognition is established in the process of constructing them. Analogies are used to establish connections between familiar and unfamiliar domains. Other techniques include making explicit the assumptions regarding the ontology of target systems, processes of idealization and de–idealization. All these processes are dependent on the process of constructing models. By themselves, computational models are too abstract to stand for things, but if correct assumptions and constraints are added, considering the knowledge of the targeted phenomenon, they can become useful representational instruments.

4. Conclusions

This research has proposed an approach to scientific representation in cognitive science which differs in various respects from the usual ways in which cognitive scientists treat representations. Representations are studied in cognitive science on the basis of theories of information processing and computational approaches. Cognitive scientists are used to thinking representations as if they were merely internal, by reducing them into mental processes. Minds process information because there are agents that represent the content of certain mental states. Thus, representation is considered as a function of a cognitive agent by means of which this agent can refer to a particular content. Then, meanings are related to the possibilities that mental contents can be associated with certain events or states of affairs in the world. This classic view of representations has been challenged by this research by considering a different type of representation that is also used in cognitive science: scientific representations.

Scientific representations are those representations involved in scientific practices – and the philosophical debate about them is particularly interested in current scientific modeling practices. Since scientific theories and concepts are too ideal for representing empirical phenomena, mediators are required when scientific knowledge is intended to represent phenomena. Scientific models are regarded as mediators between scientific knowledge and real things.

Scientific models, as stated, are not inherently representational. Indeed, although representations are used by scientists to understand phenomena, scientific techniques can be non–representational as well. Scientific representations are not the ultimate goals of scientific enterprises; they are not even always needed in them. Scientists pursue several goals that determine which techniques they employ. Pragmatically–oriented scientific practices can use non–representational techniques as far as they are not interested in depicting phenomena, but rather in manipulating data for practical concerns. In this regard, a model can be constructed to be computationally tractable, but with the cost of losing representational power, or vice versa. Considering this, the representational power of a model is related to the specific intentions of scientists.

This research has considered how pragmatic drives influence the theoretical goals of scientific enterprises. By being committed with a pragmatic approach, it differs from essentialist views of scientific representation and scientific realism. According to them, sciences explain phenomena by considering their causal dependencies and structural relations. To do this, sciences employ concepts and models that are aimed to represent these entities. In the account proposed here, scientific representations are rather seen as particular scientific achievements that are possible only if certain specific conditions are met. For instance, a scientist could create a model considering the various constraints she knows a target system has. By doing this, she would probably make the model a better depiction of certain phenomenon. This decision could, in turn, come with the cost of losing predictive power. In that sense, representational power is not something inherent to a model and it rather involves a constant negotiation between the different possibilities of models.

This research has been particularly interested in computational models in cognitive science. It is expected that it could lead to a revision of the arguments presented by their main critics (enactive approach). Radical enactivism, for instance, denies that representations play a relevant role in cognitive processes. Instead, they pretend that most (if not all) cognitive processes are originated from the primitive engagements of agents with their environments (Vernon, Lowe, Thill and Ziemke, 2015). Certainly, these engagements do not involve any form of representation. However, it cannot be inferred from the fact that lower domains do not employ representations that this happens in higher cognitive processes. Higher domains such as scientific cognition are qualitatively different from these primitive engagements. In them, representations are explicitly invoked and employed in various ways., it makes no sense to deny the role of representations in certain cognitive acts just because primary cognitive engagements do not need them.

This misunderstanding has also been reinforced by the idea that representations in higher cognitive domains are some form of mental or internal states. But scientific knowledge is not just in the minds of scientists. It can be found in external artifacts such as articles, books, etc., and a community that shares certain meaningful experiences. Reducing scientific cognition to the activities of isolated minds simply ignores this. Although this research has not developed a position regarding the internal or external nature of scientific representations, it is consistent with the claim that they are not entities existing only in scientists' minds. Scientific models are rather external and observable entities that are part of a meaningful environment, supporting certain sense–making processes.

The distinction between the two senses of representations in cognitive science (scientific techniques and a phenomenon of interest) is at the basis of the exploration this research has conducted concerning the representational techniques of cognitive scientists. Representations can be part of the scientific frameworks that are used for giving accounts phenomena (for instance, causal or mechanistic), but only if certain assumptions are met.

Section 2 introduced the recent debates on scientific representations in philosophy of science and a phenomenological approach of theoretical acts. Scientific representations have been considered as relations generated in the interaction between skilled agents that intend to use models to stand for things. According to semantic views, as discussed in section 2.1, scientific representation is a relation that involves a source, a target system and agents. Against some realist interpretations in philosophy of science, this section bets for an understanding of target systems, the presumed structures of phenomena, in terms of hypothetical systems or theoretical assumptions created in the process of constructing models rather than existing entities.

According to semantic views, scientific representations are constituted either by relations of similarity or isomorphism, but there are several arguments against these two options. Pragmatic views emphasize the material conditions and particular intentionality needed to establish representational relations. Instead of assuming that representing is an inherent property of certain models, they assert that certain specific practices are needed to make models representationally powerful. These practices were further explored in section 3 in the context of computational models in cognitive science.

Section 2.2 has provided a general description of the intentionality behind these scientific acts that create such entities, describing the constituents of theoretical acts and the general commitment of postulating the spatiotemporal existence of theoretical phenomena. According to the concept of theoretical attitude, scientists' intentions determine the possible type of knowledge they can gain and the election of techniques. The outcomes of these acts, the representational power, capacity to predict, tractability, etc., are therefore dependent on these particular attitudes.

Section 3 has described the computational theory of mind (CTM) and the computational–representational understanding of the mind (CRUM). They can be considered as particular theoretical attitudes of cognitive scientists. According to this, cognitive sciences assume that the mind is or can be described in terms of a computing machine. This thesis cannot be supported in a strong metaphysical sense. The pragmatic view only treats computational modeling as a fruitful analogy for studying the mind. It was argued that CTM and CRUM underlie connectionist approaches to cognition, while the classic cognitivist views rather endorse CTM. Finally, this section has discussed how artificial neural networks models (ANN models) can serve for representing phenomena.

Section 3 has also argued that computational models are not representational by themselves. Modelers' strategies make them representational by adding constraints and eliciting the assumptions of models' construction that leads to an understanding of the extent of the any posited representational relation. In the case of computational models, which by themselves are highly abstract or idealized, they must be de–idealized in order to represent and/or simulate phenomena in accordance with the goals of scientists. Since ANN models are these kinds of models, they need to achieve the aforementioned conditions for representing phenomena.

Section 3.2 has described the material conditions that need to be taken into consideration when ANN models are constructed for representing phenomena. The first condition is to establish an analogy between a computational model and a particular cognitive process. Analogies support the understanding of unknown, unfamiliar or uncontrollable domains by relating them with domains with the opposite characteristics. Using computational models can be a good strategy to gain understanding of cognitive phenomena as far as they make complex phenomena tractable. Other techniques involve idealizations that make models computationally tractable, supporting the discovery of mechanisms that can explain the behaviors of phenomena in causal terms and the use of multiple models. De–idealizations are also useful for making models less abstract and to resemble the phenomena they are directed at. Without this, computational models are too abstract for representing.

The constraints and assumptions that are taken in the process of constructing a model are regarded as the material conditions that enable representation. According to this view, certain models are better for representing than others. The negotiation between the abstract or concrete character of a model, the kind of analogies established, and the different goals that direct processes of idealization determine the nature of a computational model and its possibilities for representing. This view assumes that modelers have certain assumptions about the phenomena they are interested in, and they construct their models in accordance with this previous knowledge. Summing up, ANN models are abstract computational tools that can achieve representation only if certain constraints are added in their construction, the extent of the representation is stated, and they are de–idealized.

The main idea of this research is precisely that models by themselves do not represent, and that this fact can be observed in the type of models used by cognitive scientists. The limitations of this research, however, are quite salient. This research has not been intended to analyze the different of models in this discipline (many of these are not computational), nor has an exhaustive treatment of the different computational models, nor even a particular one (ANN models are too broad for this). It only purports to emphasize the general principles that operate in the representational practices that use computational models. Nevertheless, exploring the uses of computational models is in accordance with the interdisciplinary nature of cognitive sciences as far as the use of these models is transversal to its different sub–disciplines.

References

Bailer-Jones, D. (2009). *Scientific Models in Philosophy of Science*. Pittsburg, CA: University of Pittsburg Press.

Barsalou, L., Santos, A., Simmons, W. K., & Wilson, C. (2008). Language and simulation in conceptual processing. In M. De Vega, A. Glenber, & A. Graesser (Eds.), *Symbols and Embodiment*: *Debates on Meaning and Cognition* (pp. 245–283). New York, NY: Oxford University Press.

Bermúdez, J. L. (2014). *Cognitive Science. An Introduction to the Science of the Mind* (2nd ed.). Cambridge, England: Cambridge University Press.

Bickhard, M, & Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science*. *Impasse and Solution*. Amsterdam, Netherlands: Elsevier.

Bueno, O., & Colyvan, M. (2011). An inferential conception of the application of mathematics. *Noûs*, *45*(2), 345–374. https://doi.org/10.1111/j.1468-0068.2010.00772.x

Boumans, M. (1999). Built-in justification. In M. Morrison, & M. S. Morgan (Eds.), *Models as Meditators. Perspectives on Natural and Social Science* (pp. 66–96). Cambridge, England: Cambridge University Press.

Brandt, L. (2009). Subjectivity in the act of representing: The case of subjective motion and change. *Phenomenology and the Cognitive Sciences*, *8*, 573–601. https://doi.org/10.1007/s11097-009-9123-9

Casey, G., & Moran, A. (1989). The computational metaphor and cognitive psychology. *The Irish Journal of Psychology*, *10*(2), 143–161. https://doi.org/10.1080/03033910.1989.10557739

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: The MIT Press.

Chernavin, G. (2016). The process of sense–formation and fixed sense–structures: Key intuitions in the phenomenology of Edmund Husserl and Marc Richir. *Russian Studies in Philosophy*, *54*(1), 48–61. https://doi.org/10.1080/10611967.2016.1169101

Chubb, C., Dosher, B., Shiffrin, R., & Zhong–L, L. (2013). *Human information processing: Vision, memory, and attention*. Washington, D.C., WA: American Psychological Association.

Clark, A., & Lutz, R. (Eds.). (1992). *Connectionism in Context*. London, England: Springer.

Colyvan, M. (2001). The miracle of applied mathematics. *Synthese*, *127*, 265–278. https://doi.org/10.1023/A:1010309227321

Colyvan, M. (2002). Mathematics and aesthetic considerations in science. *Mind*, *111*(441), 69–74. https://doi.org/10.1093/mind/111.441.69

Da Costa, N., & French, S. (2000). Models, theories and structures: Thirty years on. *Philosophy of Science*, *67*, 116–127.

Dawson, M. (2005). *Connectionism. A Hands–On Approach*. Oxford, England: Blackwell Publishing.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense–making. An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*, 485–507. https://doi.org/10.1007/s11097-007-9076-9

Dennis, M., Simic, N., Bigler, E. D., Abildskov, T., Agostino, A., Taylor, H. G., … Yeates, K. O. (2013). Cognitive, affective, and conative theory of mind (ToM) in children with traumatic brain injury. *Developmental Cognitive Neuroscience*, *5*, 25–39. https://doi.org/10.1016/j.dcn.2012.11.006

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, *12*, 961–974. https://doi.org/10.1016/S0893-6080(99)00046-5

Egmont–Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks– a review. *Pattern Recognition*, *35*, 2279–2301. https://doi.org/10.1016/S0031-3203(01)00178-9

Farkaš, I. (2012). Indispensability of computational modeling in cognitive science. *Journal of Cognitive Science*, *13*, 401–435.

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., … Pellicano, E. (2019). Making the future together: Shaping autism research through meaningful participation. *Autism*, *23*(4), 943–953. https://doi.org/10.1177/1362361318786721

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science*, *13*(2), 103–121. https://doi.org/10.1080/02698599908573612

Frigg, R. (2006). Scientific representation and the semantic view of theories. *Theoria*, *55*, 49–65. https://doi.org/10.1387/theoria.553

Fuchs, T., & de Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense–making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*, 465–486. https://doi.org/10.1007/s11097-009-9136-4

Gallagher, S., & Zahavi, D. (2012). *The Phenomenological Mind*. *An introduction to philosophy of mind and cognitive science*. London, England, New York, NY: Routledge.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3-4), 455–479. https://doi.org/10.1080/02643290442000310

Garrido, A. (2010). Connectionism vs. computational theory of mind. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *1*(1), 40–42.

Gentner, D. (2010). Psychology in cognitive Science: 1978–2038. *Topics in Cognitive Science*, *2*, 328–344. https://doi.org/10.1111/j.1756-8765.2010.01103.x

Giere, R. (1988). *Explaining Science: A Cognitive Approach.* Chicago, IL: University of Chicago Press.

Giere, R. (Ed.) (1992). *Cognitive Models of Science*. Minneapolis, MN: University of Minnesota Press.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science, 71*, 742–752. https://doi.org/10.1086/425063

Giere, R. (2010). An agent–based conception of models and scientific representation. *Synthese*, *172*, 269–281. https://doi.org/10.1007/s11229-009-9506-z

Godfrey–Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, *21*(5), 725–740. https://doi.org/10.1007/s10539-006-9054-6

Gómez–Ramirez, J. (2014). *A New Foundation for Representation in Cognitive and Brain Science*. *Category Theory and the Hippocampus*. Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-007-7738-5

Hackenberg, T. (2009). Realism without truth: A review of Giere's science without laws and scientific perspectivism. *Journal of the Experimental Analysis of Behavior, 91*(3), 391–402. https://doi.org/10.1901/jeab.2009.91–391

Hardy, L. (1992). The idea of science in Husserl and the tradition. In L. Hardy, & L. Embree (Eds.), *Phenomenology of Natural Science* (pp. 1–34). Dordrecht, Netherlands: Springer.

Harnad. S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Humphreys, P. (2002). Computational models. *Philosophy of Science*, *69*(53), 1–11. https://doi.org/10.1086/341763

Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford, NY: Oxford University Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenological Philosophy*. Translated by D. Carr. Evanston, IL: Northwestern University Press.

Husserl, E. (1982). *Cartesian Meditations. An Introduction to Phenomenology*. Translated by D. Cairns. The Hague, Netherlands: Martinus Nijhoff Publishers.

Husserl, E. (1983). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First book: General introduction to a pure phenomenology*. Translated by F. Kersten. The Hague, Netherlands: Kluwer Academic Publishers.

Husserl, E. (1989). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second book: Studies in the Phenomenology of Constitution*. Translated by R. Rojcewicz and A. Schuwer. Dordrecht, Netherlands: Kluwer Academic Publishers.

Husserl, E. (2001a). *Logical Investigations*, *vol. 1*. Translated by J. N. Finlay. London, England, New York, NY: Routledge.

Husserl, E. (2001b). *Logical Investigations*, *vol. 2*. Translated by J. N. Finlay. London, England, New York, NY: Routledge.

Husserl, E. (2008). *Introduction to Logic and Theory of Knowledge*. *Lectures 1906/07*. Translated by C. Ortiz Hill. Dordrecht, Netherlands: Springer.

Husserl, E. (2019). *First Philosophy*. *Lectures 1923/24 and related text from the manuscripts (1920–1925)*. Translated by S. Luft and T. M. Naberhaus. Dordrecht, Netherlands: Springer.

Hutto, D, & Myin, E. (2013). *Radicalizing Enactivism: Basic minds without content*. Cambridge, MA: The MIT Press.

Jorna, R. (1990). The computational theory of mind and constrains on the notions of symbol and mental representation. In J. Baker, M. Hyland, R. Hezewijk, & S. Terwee (Eds.), *Recent Trends in Theoretical Psychology. Proceedings of the Third Biennial Conference of the International Society for Theoretical Psychology*, *April 17–21, 1989* (pp. 275–287). New York, NY: Springer–Verlag.

Kant, I. (2007). *Critique of Pure Reason, Second edition*. Translated by N. Kemp Smith. New York, NY: Palgrave Macmillan.

Kelley, T. (2003). Symbolic and sub–symbolic representations in computational models of human cognition. *Theory & Psychology*, *13*(6), 847–860. https://doi.org/10.1177/0959354303136005

Kiverstein, J., & Rietveld, E. (2018). Reconceiving representational–hungry cognition: an ecological–enactive proposal. *Adaptive Behavior*, *26*(3), 147–163. https://doi.org/10.1177/1059712318772778

Knuuttila, T. (2003). Is representation really in crisis? *Semiotica*, *143*(1–4), 95–111.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, *72*(5), 1260–1271. https://doi.org/10.1086/508124

Knuuttila, T. (2010). Some consequences of the pragmatic approach to representation. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSA Epistemology and Methodology of Science. Launch of the European Philosophy of Science Association* (pp. 139–140). Dordrecht: Springer.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model–based representation. *Studies in History and Philosophy of Science*, *42*(2), 262–271. https://doi.org/10.1016/j.shpsa.2010.11.034

Knuuttila, T. (2017). Imagination extended and embedded: Artefactual versus fictional accounts of models. *Synthese*. https://doi.org/10.1007/s11229-017-1545-2

Knuuttila, T., Honkela, T., & Rusanen, A.–M. (2007). Self–organizing maps as traveling computational templates. In *Proceedings of International Joint Conference on Neural Networks*. Orlando, FL: IEEE. https://doi.org/10.1109/IJCNN.2007.4371134

Knuuttila, T., & Loettgers, A. (2012). The productive tension. Mechanisms vs. templates in modeling the phenomena. In P. Humphreys, & C. Imbert (Eds.), *Models, Simulations and Representations* (pp. 3–24). New York, NY: Routledge.

Knuuttila, T., & Merz, M. (2009). Understanding by Modelling. An objectual approach, In de H. Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding*. *Philosophical Perspectives* (pp. 146–168.) Pittsburg, CA: Pittsburg University Press.

Knuuttila, T., & Voutilainen, A. (2003). A parser as an epistemic artifact: A material view on models. *Philosophy of Science, 70*(5), 1484–1495. https://doi.org/10.1086/377424

Lappi, O. (2007). Computational templates, neural networks dynamics, and symbolic logic. *International Joint Conference on Neural Networks* (pp. 1226–1239). New Jersey, NJ: IEEE. https://doi.org/10.1109/IJCNN.2007.4371133

Lewandowsky, S., Pilditch, T, Madsen, J., Oreskes, N., & Risbey, J. (2019). Influence and seepage: An evidence–resistant minority can affect public opinion and scientific belief formation. *Cognition*, *188*, 124–139. https://doi.org/10.1016/j.cognition.2019.01.011

Li, P., & Zhao, X. (2013). Self–organizing map models of language acquisition. *Frontiers in Psychology*, *4*(828), 1–15. https://doi.org/10.3389/fpsyg.2013.00828

Marchetti, I., Koster, E., Sonuga–Barke, E., & De Raedt, R. (2012). The default mode network and recurrent depression: A neurobiological model of cognitive risk factor. *Neuropsychology Review*, *22*, 229–251. https://doi.org/10.1007/s11065-012-9199-9

McClelland, J., Rumelhart, D., & Hinton, E. (1987). The appeal of parallel distributed processing. In J. McClelland, D. Rumelhart, & E. Hinton (Eds.), *Parallel Distributed Processing, vol. 1. Explorations in the microstructure of cognition* (pp. 3–44). Massachusetts, MA: The MIT Press.

Milkowski, M. (2013). *Exploring the Computational Mind*. Cambridge, MA: The MIT Press.

Milkowski, M. (2018). From computer metaphor to computational modeling: The evolution of computationalism. *Minds and Machines*, *28*, 515–541. https://doi.org/10.1007/s11023-018-9468-3

Morgan, A. (2014). Representations gone mental. *Synthese*, *191*(2), 213–244. https://doi.org/10.1007/s11229-013-0328-7

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. Morrison, & M. S. Morgan (Eds.), *Models as Meditators*. *Perspectives on Natural and Social Science* (pp. 10–38) Cambridge, England: Cambridge University Press.

Mormann, T. (1991). Husserl's philosophy of science and the semantic approach. *Philosophy of Science*, *58*(1), 61–83.

Negele, A., Kaufhold, J., Kallenbach, L., & Leuzinger–Bohleber, M. (2015). Childhood trauma and its relation to chronic depression in adulthood. *Depression Research and Treatment*. https://doi.org/10.1155/2015/650804

Nehaniv, C. (Ed.). (1999). *Computation for Metaphors, Analogy and Agents*. Berlin, Germany: Springer.

Neisser, U. (1967). *Cognitive Psychology*. New York, NY: Appleton–Century–Crofts.

Newell, A. (1980). Physical symbol systems. *Cognitive Science. A Multidisciplinary Journal*, *4*(2), 135–183. https://doi.org/10.1207/s15516709cog0402_2

Newton, K. (2016). *Problem–Solving*: *Strategies, Challenges and Outcomes*. New York, NY: Nova.

Northoff, G., & Heinzel, A. (2006). First–person neuroscience: A new methodological approach for linking mental and neuronal states. *Philosophy, Ethics, and Humanities in Medicine*, *1*(3), 1–10. https://doi.org/10.1186/1747-5341-1-3

Núñez, R., Allen, M., Gao, R., Miller, C., Relaford–Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behavior*, *10*. https://doi.org/10.1038/s41562-019-0626-2

Pecher, D., & Zwaan, R. (Eds.). (2010). *Grounding Cognition. The Role of Perception and Action in Memory, Language and Thinking*. New York, NY: Cambridge University Press.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, *37*, 1–38. https://doi.org/10.1007/s10867-010-9195-3

Pylyshyn, Z. (1985). *Computation and Cognition. Towards a Foundation for Cognitive Science*. Cambridge, MA: The MIT Press.

Rusanen, A.–M., & Ylikoski, P. (2007) Neural network templates and their interpretation. *Proceedings of International Joint Conference on Neural Networks*. Orlando, FL: IJCNN. https://doi.org/10.1109/IJCNN.2007.4371382

Samuels, R. (2018). Classical computational models. In M. Sprevak, & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 103–119). London, England, New York, NY: Routledge.

Schurz, M., & Perner, J. (2015). An evaluation of neurocognitive models of theory of mind. *Frontiers in Psychology*, *6*(1610). https://doi.org/10.3389/fpsyg.2015.01610

Shapiro, J. A. (2007). Bacteria are small but not stupid: Cognition, natural genetic engineering and sociobacteriology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *38*(4), 807–819. https://doi.org/10.1016/j.shpsc.2007.09.010

Sharkey, N. (1991). Connectionist representation techniques. *Artificial Intelligence Review*, *5*, 143–167. https://doi.org/10.1007/BF00143759

Shim, M. (2011). Representationalism and Husserlian phenomenology. *Husserl Studies, 27*, 197–215. https://doi.org/10.1007/s10743-011-9093-y

Sprevak, M., & Colombo, M. (Eds.). (2018). *The Routledge Handbook of the Computational Mind*. London, England, New York, NY: Routledge.

Steels, L. (2008). The symbol grounding problem has been solved, so what's next?. In M. De Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and Embodiment: Debates on meaning and cognition* (pp. 223–244). Oxford, NY: Oxford University Press.

Stewart, J. (2010). Foundational issues in enaction as a paradigm for cognitive science: From the origin of life to consciousness and writing. In J. Stewart, O. Gapenne, O., & E. Di Paolo (Eds.), *Enaction. Toward a New Paradigm for Cognitive Science* (pp. 1–32). Cambridge, MA: The MIT Press.

Stinson, C. (2018). Explanation and connectionist models. In M. Sprevak, & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 120–132). London, England, New York, NY: Routledge.

Suárez, M. (2003a). Hacking Kuhn. *Revista de Filosofía*, *28*(2), 261–284.

Suárez, M. (2003b). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*(3), 225–244. https://doi.org/10.1080/0269859032000169442

Suárez, M. (2004). An inferential conception of scientific representation, *Philosophy of Science*, *71*, 767–779. https://doi.org/10.1086/421415

Suárez, M. (2005). The semantic view, empirical adequacy and application. *Crítica*, *37*(109), 29–63.

Suárez, M. (2010). Scientific representation. *Philosophy Compass*, *5*(1), 91–101. https://doi.org/10.1111/j.1747-9991.2009.00261.x

Suárez, M. (2016). Representation in Science. In Humphreys, P. (Ed.), *The Oxford Handbook of Philosophy of Science* (pp. 440–459). Oxford, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.25

Suárez, M., & Pero, F. (2019). The representational semantic conception. *Philosophy of Sciences*, *86*, 344–365. https://doi.org/10.1086/702029

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Urbana, IL: University of Illinois Press.

Thagard, P. (2005). *Mind*. *Introduction to Cognitive Science* (2nd ed.). Cambridge, MA: The MIT Press.

Thompson, E. (2007). *Mind in Life. Biology, phenomenology and the sciences of Mind*. London, England, Boston, MA: Harvard University Press.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford, England: Oxford University Press.

Szaif, J. (2007). Doxa and epistêmê as modes of acquaintance in Republic V. *Études Platonciennes*, *4*, 253–272. https://doi.org/10.4000/etudesplatoniciennes.915

Wang, Y. (2007). The cognitive process of decision making. *International Journal of Cognitive Informatics and Natural Intelligence*, *1*(2), 73–85. https://doi.org/10.4018/jcini.2007040105

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, *104*(12), 639–659. https://doi.org/10.5840/jphil20071041240

Van Fraassen, B. (1985). Empiricism in the philosophy of science. In P. Churchland, & C. Hooker (Eds.), *Images of Science: Essays on realism and empiricism* (pp. 209–224). Chicago, IL: University of Chicago Press.

Van Fraassen, B. (2008). *Scientific Representation: Paradoxes of perspective*. Oxford, NY: Clarendon Press.

Varela, F., Thompson, E., & Rosch, E. (2016). *The Embodied Mind* (Rev. Ed.). Cambridge, MA: The MIT Press.

Velmans, M. (1999). Intersubjective science. *Journal of Consciousness Studies*, *6*(3), 299–306.

Vernon, D., Lowe, R., Thill, S., & Ziemke, T. (2015). Embodied cognition and circular causality: on the role of constitutive autonomy in the reciprocal coupling of perception and action. *Frontiers in Psychology*, *6*(1660), 1–9. https://doi.org/10.3389/fpsyg.2015.01660

Zahavi, D. (2003). *Husserl's Phenomenology*. California, CA: Stanford University Press.

Zhang, J., & Patel, V. (2006). Distributed cognition, representation, and affordance. *Pragmatics & Cognition*, *14*(2), 333–341.

Appendix A:

Abstract (in German)

Das Konzept der Repräsentation ist zentral für das Unterfangen der Kognitionswissenschaft, wo es hauptsächlich im Sinne von internen oder mentalen Zuständen betrachtet wird. Diese Sicht wird innerhalb der Wissenschaftsphilosophie neuerdings infrage gestellt. Demnach sind Repräsentationen ebenfalls Teil des Forschungsrahmens, in dem kognitive Phänomene untersucht werden. Wissenschaftliche Repräsentationen sind jene Art von Repräsentationen die in wissenschaftlicher Forschung vorkommen, welche derzeit unter Verwendung von Modellen durchgeführt wird. Modelle, ihrerseits, haben vielfältige epistemische Möglichkeiten, die nicht nur auf Repräsentationen beschränkt sind. Es wird argumentiert, dass Modelle nur nach einer Ent–idealisierungen, der Hinzufügung von Randbedingungen und der Rechtfertigung von Annahmen, für die Repräsentation von Phänomenen verwendet werden können.

In Anbetracht der mannigfaltigen Möglichkeiten von Modellen untersucht die vorliegende Arbeit folgende Frage: Wie werden wissenschaftliche Repräsentation in der Kognitionswissenschaft erreicht? Die Arbeit diskutiert die wichtigsten aktuellen wissenschaftlichen Beiträge aus Wissenschaftsphilosophie und phänomenologischen Ansätzen zum Akt des Repräsentierens, um eine allgemein Interpretation von wissenschaftlichen Repräsentationen zu skizzieren. Anschließend werden Prozesse diskutiert welche relevant sind für die Konstruktion von Modellen die zur Repräsentation bestimmt sind. Das sind: die Idealisierung, die Ent–idealisierung, der Analogieschluss, des Hinzufügung von Randbedingungen und die Explizitmachung von Annahmen. Diese Betrachtungen werden Unterstützt von einer Vielzahl von Beispielen der Verwendung von Computermodellen in der Kognitionswissenschaft, welche zeigen sollen wie sich diese theoretischen Ansätze zu aktuellen wissenschaftlichen Praktiken verhalten. Es wird auch diskutiert wie Repräsentationen in zwei der wichtigsten Paradigmen des Feldes (Kognitivismus und Konnektionismus) verstanden werden.

Ziel dieser Masterarbeit ist die Aufstellung eines philosophisches Arguments mit Relevanz für die Epistemologie der Kognitionswissenschaft. Die Vorgehensweise ist klar interdisziplinär, da es eine Art von Modellierung untersucht die in einer Vielzahl von Teilgebieten der Kognitionswissenschaft verwendet wird.