U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Foundation of a Hierarchical Stacked Neural Network model for Simulating Cognitive Development

Doctoral Program in Biomedical Engineering

Academic Year 2017/2018

Authorship:

Sofia Ribeirinho Soares Ferreira Leite

Clinical Psychologist at Avicena, Clínica Psiquiátrica, Porto, Portugal
Researcher at AI4Health group, CINTESIS, University of Medicine, Porto, Portugal
Researcher at Dare Association, Inc., Cambridge, MA, USA

Supervision:

Professor Pedro Pereira Rodrigues

Principal Investigator at CINTESIS, University of Medicine, Porto Portugal
Researcher at AI4Health group, CINTESIS, University of Medicine, Porto, Portugal
Assistant professor at Department of Community Medicine, Information and Health Decision Sciences, University of Medicine, Porto, Portugal
Principal Investigator at INESC TEC

Professor Michael Lamport Commons

Corresponding Member at Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Harvard University, MA, USA
Head of Research at Dare Association, Inc, Cambridge, MA, USA

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

Thank you,

Professor Michael Lamport Commons and Professor Patrice Marie Miller, (Michael and Patti), for everything that is to mention and for everything else that is unmentionable. Professor Pedro Pereira Rodrigues, for your eyebrows moving whenever you didn't agree, for the way you let things be, and for your precision. Professor José Carlos Príncipe, for the beginning of this amazing journey, for your thoughts, and for tremendously helpful conversations. Professor Fernando Jorge Monteiro, for your undoubtful presence and trust. Professor João Marques Teixeira, for your wisdom and friendship. Professor Rui Ribeiro, for having let me learning calculus, which is different from having taught me. Professor James McClelland, for the extensive and dedicated review of Ch. II-B, and for the funded participation at the NCPS'15 conference. Fulbright Commission, for an absolutely beautiful, rich and life-changing opportunity of studying in the USA, Harvard University, for a period of six months with a Fulbright Research Grant. Professor Carlos Mota Cardoso, for participating in so many facets of this work and journey, more than I thought it could possibly have. Professor António Ramos Silva, for your straightforward and acute attention to detail, and for always making things possible and simple. My dear family and my dear friends, for celebrating the good news. Thank you, António. Thank you, Maria.

Thank you.

ABSTRACT

ABSTRACT

Background: According to developmental psychology, individuals develop throughout stages of development. At each stage, they are able to solve an increasing number of problems and increasingly demanding problems. Artificial Intelligence (AI) aims to transcribe human information processing capabilities to the machine, eventually providing machines with procedures that allow for mimicking adaptation. This means learning in a developmental way, which implies that learning and development are discriminated and coordinated in a model of human information processing.

Problem: Connectionist models are a class of AI models that approximate the general laws of natural information processing in the central nervous system. They are strong learners, yet, they do not learn in a developmental way. First, learning in a connectionist model is a continuous process, based on gradient descent techniques of weight update; second, each new activation pattern of a network substitutes (or updates) the previous one. This contradicts the discrete and cumulative nature of development. The fact that development is not discriminated and complemented with learning in a connectionist model is at the basis of the flexibility/stability dilemma.

Aim: This dissertation aims to build a developing connectionist model, built in stacks. Stacks are a synonym of stages of development. It was the goal of the present work to provide a method that identifies the structure of each stack composing the global model (stage of development), to identify what changes from one stack to the immediately next stack (stack transition), and to extract a progression of change throughout stacks (developmental progression).

Method: The method here designed results from the overlap of three communicating disciplines — Developmental Psychology, Cognitive Neuroscience, and Artificial Intelligence. The first thing to do to build a model that learns in a development way is to define stage of development. The Model of Hierarchical Complexity (MHC) was adopted as the theoretical reference, since it defines stages of development as an Order of

Hierarchical Complexity (OHC). The OHC is the factor that ascribes stability to the performance within stage. The second thing is to represent stages of development in a model of cognitive development. The present method has a two-folded application. It can be applied to the field of developmental cognitive neuroscience, in order to identify how stages of development are represented in the brain, and to connectionist models, in order to identify the minimal complexity network structure that represents each stage of development. Network structures were evaluated based on a varying number of units, layers, and connectivity pattern among units. This rationale relied on assumptions from Complex System Theory (CST), which gives a perspective over brain functioning as mainly implemented in a network of internal dependencies, as much as in connectionist models. By comparing networks structures of adjacent OHC, the changes that undergo from one structure to the next can be determined and used to ascribe developmental properties to the model. The simulation scenario was the balance scale test, a developmental test applied to children that evaluates their current stage of development. This scenario has been object of interest since the 80's. Yet, existing attempts fall on limitations attached to the a priori definition and representation of stages and stage transitions.

Results: Results allowed for identifying a connectionist structure underlying each OHC, which is represented by the complexity of operations and the number of problem dimensions. They also allowed for identifying two types of stack transition — memory-based and operationally-based transitions. Memory-based transitions occur when the problem dimensions increase; operationally-based transitions occur when the operation increases in complexity. Operationally-based transitions seems to underlie the transition from abstract to formal reasoning, associated with higher-order cognition and apparently only present in humans. When the connectivity pattern was allowed to vary, there was a tendency for independently trained networks converging to the same number of connections. Another important result was that networks with input connectivity patterns showed improved performance, decreased structural complexity, and greater tendency for biological plausibility. Finally, results allowed for identifying a developmental progression across network structures, even if departing from different structures. Interestingly, this was only true for structures sharing the same connectivity pattern.

Discussion: The set of studies composing this dissertation allow for stating that the "Foundation of a Hierarchical Stacked Neural Network model for Simulating Cognitive Development" is based on a system that is built in stacks, where 1) each stack is triggered by the OHC of the problem to solve, 2) each next stack can be built out of the elements that compose the previous stack, which suggests the term "Structural Integration" for describing the stack transition mechanism, and 3) each previous stack is protected by the interference of the next stack, as its components remain triggered by the OHC of the problem to solve. Results are preliminary, though. They are most useful to corroborate the plausibility of the proposed method, as a way for approaching the problem of flexibility and stability in artificial learning models, especially hierarchical stability and developmental processes.

Contributions: This dissertation mainly contributed for the definition of theoretical and methodological guidelines, here corroborated by a set of studies, that led to approaching a model of cognitive development based on the MHC. These theoretical and methodological guidelines can likely contribute to a new research line that triangulates the fields of developmental psychology, developmental cognitive neuroscience, and AI.

RESUMO

RESUMO

Introdução: Segundo a psicologia do desenvolvimento, os indivíduos desenvolvem-se por estadios. A cada estadio, mais problemas e problemas mais complexos são resolvidos. A Inteligência Artificial (IA) procura transcrever as propriedades do processamento de informação para a máquina, eventualmente munindo a máquina de procedimentos que simulem o processo adaptativo. Ou seja, procedimentos que habilitem a máquina a aprender de forma desenvolvimental, o que implica a discriminação e a coordenação dos processos de aprendizagem e de desenvolvimento.

Problema: Os modelos conecionistas formam uma classe de modelos de IA que pretende simular as leis gerais associadas ao processamento de informação no sistema nervoso central. São modelos com fortes capacidades de aprendizagem, mas não aprendem a resolver problemas de forma desenvolvimental. Primeiro, a aprendizagem nestes modelos é um processo contínuo, baseado em técnicas de descida de gradiente para atualização dos pesos que conduzem à minimização do erro; segundo, cada novo padrão de ativação (padrão composto pelos pesos atualizados) substitui o anterior. Estes dois aspetos contradizem a natureza descontínua e cumulativa do processo desenvolvimental. O facto de o processo desenvolvimental não se encontrar discriminado e em complementaridade com o processo de aprendizagem está na base do chamado "dilema entre flexibilidade e estabilidade" do sistema.

Objetivo: Este trabalho procura construir um modelo com uma estrutura em andares, em que cada andar corresponde a um modelo conecionista, ou em rede, e representa um estádio desenvolvimental. Foi objetivo da presente dissertação apresentar um método que identificasse a estrutura particular de cada andar (estádio de desenvolvimento), que identificasse, também, as alterações que se aplicam na transição de um andar para o que se lhe segue (transições entre estádios) e que permitisse, ainda, extrair uma progressão de alterações ao longo de todos os andares (progressão desenvolvimental).

Método: O método aqui definido resulta da interação de três disciplinas próximas — psicologia do desenvolvimento, neurociência cognitiva e inteligência artificial. O primeiro requisito para construir um modelo que aprenda de forma desenvolvimental consiste na definição de estádio de desenvolvimento. Neste sentido, o Modelo de Complexidade Hierárquica (MCH) foi adotado como referência teórica, uma vez que define estádio de desenvolvimento em função da Ordem de Complexidade Hierárquica do problema (OCH). A OCH é o fator que atribui estabilidade de desempenho intra-estádio. O segundo requisito consiste na delineação de um método que permita representar estádios de desenvolvimento num modelo de desenvolvimento cognitivo. Foi delineado um método com aplicação bipartida. Pode ser aplicado à disciplina de neurociência cognitiva desenvolvimental, permitindo identificar a representação neuroanatómica e neurofuncional dos estádios de desenvolvimento, e à disciplina de IA em modelos conecionistas, permitindo identificar a estrutura de mínima complexidade que consegue representar o processamento de informação correspondente a cada estádio, ou andar. A estrutura de cada rede, ou de cada andar, foi avaliada com base num variável número de unidades, camadas de unidades e padrão de conectividade entre unidades. Este racional baseou-se em premissas extraídas da Teoria dos Sistemas Complexos (TSC), que compreendem a perspetiva do funcionamento cerebral como uma rede de dependências internas, tanto quanto nos modelos conecionistas. Ao comparar a estrutura de andares que representem OCH sucessivas, podem determinar-se as diferenças necessárias entre a estrutura de um andar e a do seu sucessivo, transcrevendo propriedades desenvolvimentais para o modelo. O cenário de simulação utilizado foi o teste da balança, um teste desenvolvimental administrado a crianças para avaliar o seu estádio de desenvolvimento. Este teste tem sido objeto de interesse para simulação desde os anos 80. No entanto, as simulações existentes pecam por limitações associadas à definição e representação de estádio de desenvolvimento e transições entre estádios.

Resultados: Os resultados obtidos permitiram identificar uma estrutura subjacente a cada OCH, ou a cada andar, que é por sua vez impactada pela complexidade das operações conduzidas e pelo número de dimensões associadas ao problema. Permitiram, também, identificar dois tipos de transições: as baseadas na memória e as baseadas nas operações. As primeiras ocorrem quando há um aumento na dimensionalidade do problema de andar para andar; as segundas, quando a operação aumenta de complexidade, parecendo, estas, estar associadas à transição de estádio das operações abstratas para as operações formais e pós-formais. Por usa vez, esta transição parece estar subjacente à cognição de ordem-superior, aparentemente única da espécie humana. Quando o padrão de conetividade entre as unidades foi, também, variável, demarcou-se uma tendência para que redes treinadas independentemente para o mesmo andar convergissem para o mesmo número de ligações entre unidades. Também importante foi o facto de modelos com um padrão de conectividade mais denso no que respeita à camada dos dados de entrada terem atingido um melhor desempenho, uma menor complexidade estrutural e uma maior tendência para cumprir plausibilidade biológica. Por último, os resultados obtidos permitiram identificar uma progressão desenvolvimental ao longo de todos os andares, mesmo que partindo de estruturas iniciais diferentes, mas desde que as estruturas sucessivas partilhassem o mesmo padrão de conetividade das anteriores.

Discussão: A sequência de experimentos que compõe esta dissertação permitiu destacar que a fundação de um modelo conecionista para simulação do desenvolvimento cognitivo deve ser baseado numa estrutura-mãe em andares, em que 1) cada andar é despoletado pela OCH do problema a resolver, 2) cada andar seguinte pode ser contruído a partir dos elementos do andar corrente, sugerindo um mecanismo de "Integração Estrutural" para descrever transições desenvolvimentais, e 3) cada andar anterior se mantém protegido da interferência de andares posteriores, na medida em que a sua ativação continua a ser despoletada pela OCH do problema a resolver. No entanto, é de salientar que os resultados obtidos são, ainda, preliminares. Acima de tudo, são úteis para corroborar a plausibilidade do método proposto para responder ao dilema entre flexibilidade e estabilidade de um sistema de IA, especialmente no que respeita à sua estabilidade hierárquica e propriedades desenvolvimentais.

Contribuições: A presente dissertação contribuiu para a delineação de premissas teóricas e metodológicas, corroboradas por uma sequência de experimentos que conduziram à iniciação de um modelo de desenvolvimento cognitivo baseado no MCH. Pretende-se que estas premissas teóricas e metodológicas contribuam, por seu lado, para um novo caminho de investigação que triangule as disciplinas de psicologia do desenvolvimento, neurociência cognitiva desenvolvimental e IA.

CONTENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AI — Artificial Intelligence

CST — Complex Systems Theory

IS — Interactive Specialization

MHC — Model of Hierarchical Complexity

MNN — Modular Neural Networks

MRI — Magnetic Resonance Imaging

OHC — Order of Hierarchical Complexity

RAM — Rule Assessment Methodology

EEG — Electro-Encephalography

SI — Saturation Index

(…) the design for a bird might be as simple as ‘‘Take a planet with some carbon and oxygen; irradiate it with sunshine and cosmic rays; and leave it alone for a few hundred million years’’ (p. 52). But the mechanism responsible for evolution is difficult to directly observe in action, and it does not appear to apply straightforwardly to a chess-playing machine. If evolution is able to produce systems that exhibit more information than is contained in their design, and information cannot be spontaneously generated, where did this extra information come from?

(The Mechanical Mind (Ch. 7))

CHAPTER I

DISSERTATION OVERVIEW

The present dissertation, entitled "Foundation of a Hierarchical Stacked Neural Network model for Simulating Cognitive Development", focuses on the transduction of the main premises of cognitive development into a connectionist model. One of these premises is that the model is built in stacks, as a synonym of stages of development. The definition of *stack* and the *transition* from one stack to another will be object of experimentation. The present overview includes a brief exposure of the relevance of the mentioned topic to Artificial Intelligence, how it has been here operationalized into objectives and experimental methods, and the contributions of the obtained results to the scientific literature. Finally, a description of the organization of the work is provided.

INTRODUCTORY NOTE

This work is conducted under the realm of **Biomedical Engineering**. Today, Biomedical Engineering designates a wide field of research dedicated to apply engineering to biology, specifically oriented to improve health and well-being. Yet, if we return to the original meaning of this expression, “Biomedical Engineering” means “the best way to produce life”.

Biomedical Engineering is an expression composed of three root words. **Bio** is a prefix meaning “life”, in Greek. **Medical** comes from the Indo-European word “MED”, which means “to evaluate, to measure”. “MED” influenced the formation of the Latin word “Mederis”, which originally meant “to know the best way to” through evaluation and measurement. However, “Mederis” became mostly used to designate those who “knew the best way to treat or cure people”, giving prevalence to the complement in detriment to the predicate. “Mederis” later influenced the formation of the term “medicus” and “medical”. These words are nowadays only associated with health and well-being. **Engineering** comes from the Latin word “Ingenius”, meaning “talent”. This word is composed of the prefix “in” added to the Indo-European root word “GEN”, which means “to create, to produce”. Thus, “Engineering” means “the art of producing” (Mota-Cardoso, 2017, personal communication).

In the present work, two of these three root components are determined. **Engineering** concerns a computational production, under the connectionist paradigm. **Bio**, or life, is assumed as the existence of change within a system as interactions with the environment proceed. Within the scope of the present work, Bio is restricted to the cognitive apparatus and how it successfully responds to problems posed by the environment. The third component — **Medical** — “the best way to” — will be exposed as the work proceeds. It reveals the importance of the method (observations and measurements) applied to reach a reliable computational production of the cognitive system, which changes as an organism develops.

MOTIVATION

Organisms of all species perceive salient stimuli and operate with such stimuli according to their biological substrate and maturational forces. As a result, they generate actions that interact with and manipulate the environment. This allows them to solve more or less problems, and more or less demanding problems, where problems are environmental situations that the organism needs to cope with, through actions, in order to adapt. According to developmental psychology, organisms develop in stages. At each next stage, they become able to solve an increasing number of problems and increasingly demanding problems. Humans are known to achieve the highest stage, as compared to other known species, opening up the way for what is called higher-order cognition. In other words, humans are able to cope with more, and more cognitively demanding problems in a flexible way.

> In terms of the complexity of actions and problems, inevitably, organisms go from performing simple actions and solving simple problems to performing more complex actions and solving more complex problems.

Lato sensu, the main goal of Artificial Intelligence (AI) is the transcription of human information processing capabilities to the machine, eventually providing machines with procedures that allow for responding to and interacting with a complex and ever-changing environment. An intelligent system is, thus, intended to reproduce higher-order cognition and to establish a rich repertoire of actions with its surroundings. Given the above, it certainly benefits from developing through the course of its "maturation", generating simple actions before more complex ones, and solving simpler problems before more demanding ones.

But what are we talking about when we talk about the "maturation" of an artificial system? In order to answer this question, this dissertation lies at the overlap between three communicating disciplines: developmental psychology, developmental cognitive neuroscience, and connectionism.

Developmental behavioral theories and developmental cognitive neuroscience inform about the maturation of individuals. Thus, they can greatly inform about the maturation of an artificial system, as well. It is argued that the observed behavioral stages of development (and transitions across stages) that organisms go through during their developing lifetime, as well as the biological changes undergoing along, need to be modelled and implemented in an algorithm, such that the algorithm approximates how humans solve sets of increasingly demanding problems.

The idea of creating a system that *develops* fosters the discrimination between learning and development.

Two major aspects differentiate these two processes. The first is that learning is substitutive, development is cumulative. Learning concerns the substitution of old, less adaptive behavioral patterns by new and more adaptive ones. Differently, a developing organism generates new stages, but maintains the ability to move down to more elementary levels of information processing, if the context so requires. More elementary levels, or stages, does not mean less adaptive stages. It means less complex stages that allow for dealing with less complex problems. Hence, development is cumulative, imbedding the capacity to move up to complex levels and down to simple levels, which provides great flexibility and adaptability. The second distinction is that learning is continuous, implying quantitative fine-tuning changes, whereas development is discontinuous, implying qualitative changes in the functional pattern at use. Given that both learning and development occur in parallel as a cause and an effect of biological maturation, it is important that AI system also differentiates between learning and development in its information processing procedures.

Connectionist models are a class of AI models, which basic components (computational units and connections) model the basic components of the natural neural architecture (neurons, roughly composed of cell bodies and axons). Connectionist models are strong artificial learners for two main reasons. First, their structure allows for distributed information processing capabilities, as the information spreads out to the computational units in the network, through connections. Second, learning is modeled by *continuously*, where poor solutions are progressively *substituted* by more adaptive ones. Input-output mappings are formed by a composite continuous differentiable function, where differentiability allows for continually updating the function parameters. When differentiability and continuity in connectionism started to produce robust results, the association between developmental psychology and the connectionist framework became more salient. Since late 80's, these models have been used at the service of developmental computational cognition, with the aim of reproducing cognitive development. Yet, there are two limiting problems in simulating development in a connectionist model.

> Development is a discontinuous process, which threatens to corrupt continuity and differentiability. If a developing procedure is employed that does not corrupt continuity and differentiability, such as in some generative architectures, it corrupts the second main premise: lower stages (or previous structures) are no longer available to the system, as they get substituted by higher stages (or new structures).

Then, how to coordinate learning and development in a connectionist AI system?

OBJECT

The present dissertation focuses on the coordination of learning and development in a connectionist model, proposing a method for studying the representation of stage transitions.

Specifically, the object of study is stage transitions, or stack transitions, employed at the service of simulating cognitive development in a *stacked* connectionist model.

According to the present perspective, a stacked neural network model is one that *develops* by stacks, where each *stack* is the computational synonym of *stage of development*. In order to study stack transitions, the present study first focuses on what a stack is, and, second, on the relation between adjacent stacks. In order to approach such study, the present dissertation uses developmental theories (behavioral and biological) as the theoretical background, principles from Complex System's Theory (CST) as methodological background, and the mathematical properties of Connectionist models as the experimental background.

The Model of Hierarchical Complexity (MHC) is a general stage theory that will be chosen as the theoretical background for a number of reasons later exposed. Developmental cognitive neuroscience will be used to provide brain-based experimental data that corroborates the existence of stages of performance at a biological level, giving some insights into how stages are imprinted in the neural architecture.

Principles from CST will be used to provide insights into how the natural language can be transduced into a computational language of the connectionist type. Complex Systems studies systems, independently of the nature of their constituents. It is used as the methodological bridge between developmental theory and data and experiments with connectionist simulations. Because CST studies the properties of systems as systems evolve, comparisons between the natural and the artificial systems based on the observed properties in action will be established, as development occurs. Hence, the use of complex systems allows for interpreting the results back and forth from the cognitive domain to the computational domain. Otherwise, the comparison between the two systems — natural and artificial — would fall in an epistemological gap.

Connectionist models are chosen for their biological plausibility in simulating the learning process, as well as for simulating the basic arrangement principles of the central nervous system: a network composed of units and connections, which activate in the face of salient information.

This class of models will be used for conducting simulations of a specific developmental learning scenario: the balance scale test. This test was developed within the domain of developmental psychology and is one of the first scenarios used to simulate development in connectionist models. Although the present work introduces a new method for cognitive development simulation, it gives continuity to a line of studies with the balance scale test that began in the late 80's.

MAIN HYPOTHESIS

According to developmental theories, stage transitions concern the moment when an individual abandons the previous functional pattern, while the adoption of a qualitatively different one is taking place. This implies a *process of emergence* of each qualitatively different functional pattern. According to a CST, a *process* is, by definition, a set of procedures that are not visible at the periphery of the system, but contained in its structure.

> The hypothesis here elaborated is that the process underlying stage transitions, or stack transitions, can be seen at the level of the structure of a developing system, both natural and artificial.

If one conducts the study of how a system's structure changes as higher stages are represented in stacks, one is closer to understanding how a system's structure encodes development, leading way to building a connectionist structure that not only *learns*, but also *develops*.

Research Question, Goal and Objectives

The study and implementation of stack transitions in a connectionist model raised the following research question: **When a developmental transition occurs, the system changes from what to what, and how?** Answering this question is the goal of the present dissertation.

In order to accomplish this goal, the following objectives were delineated:

1. **Validation of Hierarchical Integration as the known cognitive mechanism underlying stage transitions.** Hierarchical Integration is the cognitive mechanism that explains stage transitions from a behavioral developmental perspective. This mechanism assumes that each higher stage is formed out of the outputs generated at the immediately preceding stage.

2. **Definition of a method for identifying structural changes in a developing connectionist system.** The first objective concerns the definition of a method for transducing stages of development into computational developing stacks. This corresponds to the transduction of transitions from the cognitive developmental domain into the computational domain, with biological plausibility. This method also aims to give some freedom for the model to represent what it needs to represent — stages of development and stage transitions — imposing the least constraints.

3. **Representation of stages in a connectionist model.** In order to simulate stage transitions, it is necessary to understand how a connectionist model represents stages. The influence of several factors, such as problem dimensions, operations that link inputs to outputs, and the number of outputs to be generated is under

question in the definition of stage of development and stage transitions in both fields.

4. **Modelling structural changes across connectionist stacks.** In order to study stage transitions, or stack transitions, it is necessary to compare networks that represent adjacent stages. The number of layers, number of units per layer, and the connectivity pattern among units was compared. Afterwards, the identification of a structural progression from the concrete stage to the systematic stage was possible.

CONTRIBUTIONS

The present dissertation provides scientific contributions at a theoretical, methodological, and experimental levels, concerning the topic under investigation: simulating cognitive development in a connectionist model. Furthermore, it provides a parallel contribution concerning the applicability of the present rationale to different fields of study and practice: evolutionary psychology and educational practices.

Specifically, at a theoretical level, the work here elaborated allows for revising some important aspects related to development. Namely, 1) how the concept of development is addressed across the cognitive and the computational fields, including the definition and inter-relation of other associated concepts, namely, complexity, learning, stability, and flexibility; 2) how stage transitions are understood, defined, and modeled; 3) how the particular Model of Hierarchical Complexity formalizes stages of development and stage transitions, which has been found to lack some clarity. All these theoretical aspects converge to strengthen the understanding of the developmental process in natural and artificial systems.

At a methodological level, this dissertation proposed a method that is valid for two intertwined fields of study, namely computational cognition and developmental cognitive neuroscience. Usually, a research method is exclusive of a determined field of studies. Yet, if the same method is conducted cross-disciplinarily, the interpretation of results is more robust and more constrained. Consequently, the theoretical constructions that such method provide are afforded with greater validity, and interpretations of results are possible in a multi-directional way. This is relevant specifically for two disciplines that communicate so tightly, and hopefully opens up the possibility that this method is put in practice beyond the scope of this work.

At an experimental level, results first allowed to corroborate the proposed method, both in what refers to represent stacks, to represent stack transitions, and to identify a structural progression across adjacent stacks.

A generative architecture based on *structural integration* was proposed, where each new structure is fine-tuned by a process here called *pattern-wise learning*. Different structural progressions are possible, but none could yet be chosen. Further work is necessary. In sum, the obtained results allowed for coordinating learning and development in a connectionist model, as long as the model is built in hierarchical stacks — a Hierarchical Stacked Neural Network model.

Finally, during the course of this dissertation, the hypothesis that stacks (or stages) exist for different staged-problems allowed for expanding its applicability for other domains: evolutionary psychology and educational practices. First, parallel work was conducted within the scope of evolutionary psychology, establishing a parallelism between orders of complexity at a developmental scale and at an evolutionary scale. Second, within the scope of education, research is being conducted in order to define a programming course based on the hierarchical complexity of the programming tasks to be learned.

AWARDS AND PUBLICATIONS

This dissertation granted two scientific awards:

- Fulbright Grant, from May to November, 2016, at Harvard University
- Funded participation by Funded by the W. K. & K. W. Estes Fund, Google DeepMind and the Rumelhart Emergent Cognitive Functions Fund in the 15th Neural Computation and Psychology Workshop, held at Drexel University, Philadelphia, PA, USA, in 2016

And five publications:

- Rodrigo Duran, Juha Sorva, and So a Leite. (2018). Towards an Analysis of Program Complexity From a Cognitive Perspective. In ICER '18: 2018 International Computing Education Research Conference, August 13–15, Espoo, Finland. ACM, New York, NY, USA, 10 pages. DOI: 10.1145/ 3230977.3230986
- Leite, S., Rodrigues, P., (2018). Simulating Developmental Cognition: Learning by Order of Complexity in Modular Stacked Neural Networks. Oral presentation. 51st Annual Meeting of the Society for Mathematical Psychology, MathPsych Conference, July 21-24, Wisconsin, USA.
- Leite, S., (2016). Successes in Cultural Evolution Raises the Variability in Humans' Highest Stage Attained. *Behavioral Development Bulletin 21*(2), 165–175. DOI: 10.1037/bdb0000033
- Leite, S., Barker, C.D., Lucas, M.G., (2016). Neural Correlates of Postformal Stages of Reasoning: Biological Determinants of Developmental Stage. *Behavioral Development Bulletin 21*(1), 33–43. DOI: 10.1037/bdb0000012
- Leite, S., Commons, M. L., Rodrigues, P.P. (2015). Primary-Stage responses to balance-scale tasks simulated in a Hierarchical Stacked Neural Network model. Poster. 48th Annual Meeting of the Society for Mathematical Psychology, MathPsych Conference, July 17-21, California, USA.

ORGANIZATION OF THE WORK

This work contains four more chapters. Chapter II is dedicated to theoretical considerations, which set the premises for elaborating a methodological approach. It is composed of four sections. Section A shows the relevance of studying and ascribing development to a connectionist model, based on the concepts of flexibility and stability. Section B identifies the factor underlying hierarchical stability of the system. In other words, this is the factor that accounts for performance invariance at each hierarchical stage of development — the Order of Hierarchical Complexity. Section C corresponds to the first objective and tests a commonly known, widely proposed, but poorly tested, mechanism for stage transition — the mechanism of Hierarchical Integration. Section D, once this mechanism has been invalidated, reviews the parameters that underlie the representation of stages in the brain. The idea is to set the ground for the delineation of a method for representing stages in a connectionist model. Chapter III responds to the second objective of the present dissertation and delineates such method. Chapter IV applies the proposed method to a set of four experiments that aim to answer the third and fourth objectives. Finally, Chapter V presents some concluding remarks that summarize what has been here conducted, including its limitations, advantages and proposals for future work.

CHAPTER II

THEORETICAL CONSIDERATIONS FOR SIMULATING COGNITIVE DEVELOPMENT IN AN ARTIFICIAL MODEL

The present chapter is composed of a preliminary section, highlighting major concepts and findings, and of four sections. Section A "Artificial Intelligence: Five Underlying Concepts and their Inter-Relation" comprehends a historical and scientific analytical overview of the problem of simulating the *mind* in a machine, highlighting the importance of using and simulating the notion of development. It pretends to clarify and inter-relate five grounding concepts, where development is included, which will be used throughout the entire work, and will hopefully hereafter find a clear(er) conceptual basis. Section B "Bridging Developmental Psychology and Computational Cognition: The Importance of a Domain-General Stage Theory" consists of an analytical review of what has been done so far concerning the simulation of cognitive development in connectionist models. Much of existing work, including this dissertation, simulates a particular scenario: the developmental balance scale test. For this reason, this is the simulation scenario covered in this analytical review. It identifies some theoretical limitations of previous work and suggests some requirements for further work. These requirements zoom in into the notion of development and deal specifically with the definition of stages and stage transitions. Section C “Developmental Cognition in Modular Neural Networks: Stage Transitions are not explained by Hierarchical Integration” is experimental. It uses the definition of stage transitions shared by behavioral developmental theories and applies it to a simple simulation, showing that developmental transitions need to be factorized, analyzed, and implemented differently. Finally, Section D “Neural Correlates of Postformal Stages of Reasoning: Biological Determinants of Developmental Stage” shows that there is still a lack of knowledge of how stages of development and stage transitions are represented in the brain. As a result, it proposes a new method that will allow for identifying how development might be imprinted in the neural signature of individuals. This method is shown to have applicability in the field of Artificial Intelligence and Computational Cognition, too, as it provides for identifying how orders of hierarchical complexity might be imprinted in connectionist stacks. Principles from Complex System’s Theory are here used to determine the main premises of the proposed method and to allow for an interpretative parallelism between both fields of study: developmental psychology and AI.

Highlights of Chapter II

- Learning and development allow for discriminating two types of flexibility and stability: horizontal and hierarchical
- A natural learning system forms representations of the environment in a developmental way (by stages of development), following the direction of increasing complexity, which corresponds to hierarchical flexibility and stability
- Hierarchical flexibility and stability bring about an additional difficulty in connectionist models — the difficulty of coordinating the continuous nature of learning and the discontinuous nature of development
- The Model of Hierarchical Complexity is the suggested theoretical adoption for simulating hierarchical flexibility and stability because it provides evidence of a structural, universal and systematic factor underlying the formation of increasingly complex representations — the Order of Hierarchical Complexity
- Developmental behavioral theories, including the MHC, have formalized stage transitions as a process of Hierarchical Integration. During this process, lower-order outputs are coordinated to form higher-order outputs
- The study here conducted rather showed that the increase in complexity lies at the operations performed by the algorithm, in interaction with the number of problem dimensions
- It lasts to determine how is it that OHC are represented in connectionist models
- A method has been proposed to identify how OHC are represented in the brain. The same method has applicability to determine how OHC can be represented in connectionist models
- That leads way to simulate cognitive development and hierarchical flexibility and stability

Section A

Artificial Intelligence: Five Underlying Concepts and their Inter-Relation

In the second half of the twentieth century, a mechanistic view of processes and functions was dominating the philosophical and scientific streams of thought, in which the understanding of the mind played a major relevance. This mechanistic view resulted from a multidisciplinary confluence — philosophy, psychology, physiology, mathematics, engineering — and ultimately coined a new field of research: Artificial Intelligence (AI). *Lato sensu*, AI derives from a certain number of operationalizable concepts, which describe a certain portion or characteristic of human behavior. Its main goal is to transcribe the properties of human information processing capabilities to the machine, eventually providing machines with procedures that respond to the environment, solving sets of highly complex problems (Pennachin and Goertzel, 2007). To this end, AI can be used both as a method for testing models of information processing and as the implementation of what is (provisionally) known about information processing in a living organism (Cassimatis, 2012). Different approaches have been experimented with varying degree of success, both theoretically and technically.

However, this diversity has flown into a progressively sparser definition the concepts AI uses and how they are operationalized.

The objective of the present study is 1) to identify some grounding concepts of AI, 2) to clarify the definition of those concepts across fields, and 3) to inter-relate those concepts with each other and across fields. For that, this work starts with a brief exposure of the historical and scientific underpinnings of AI from the perspective of psychology. The concepts of flexibility, stability, learning, development, and complexity will be identified as major instances of cognition and adaptation. A solution for inter-relate these concepts will be proposed. Two types of flexibility and stability will be identified, which can be interpreted as horizontal and hierarchical. These two types correlate with a representation of learning and development, respectively, where development, which is hierarchical, follows the direction of increasing complexity. As will be exposed, this solution has biological and computational plausibility. Second, the implementation of these concepts in the particular case of connectionist models will be reviewed. Connectionist models are chosen as the class of artificial learning models of reference because of their rough resemblance of the central nervous system, as well as for their known strong computational ability to learn highly complex functions (Rojas, 1996). Yet, it will be shown they fall short on reproducing both types of flexibility, specifically hierarchical flexibility (French, 1999; McCloskey and Cohen, 1989).

It is beneficial that general theoretical and methodological concepts and principles are defined, so as to underlie the object, objective and main premises of AI. That way, several branches and layers of intelligence, adaptation, and problem-solving can be joined together as a common computational entity (Johnson, 2011). This allows for scaling up the work that is conducted at smaller scales at different specialized laboratories. Also, this does not compromise the necessity of their specialization, but simultaneously does not compromise the possibility of integrating these specialized attempts into a global framework.

A shared definition of fundamental concepts and methods certainly allows for raising the *qualia* of AI as a field running over its own rules instead of running over the rules of the fields which created it.

This eventually renders AI to become a cross-paradigm, rather than a set of overlapping paradigms.

1. Underpinnings of AI from the perspective of Psychology

Back to the late eighteenth century, David Hume seeded the idea of a unitary factor contracting the complexity of the mind, elaborating a Representational Theory of Mind in “Treatise of the Human Nature” (Hume, 1888). Inspired by Newton, who reduced the explanation of movement to a unitary force, Hume was seemingly driven by the goal of finding the same force that would prove to be the engine of the mind. Despite the critics on the incompleteness of his work, Hume is first credited for re-introducing the experimental method in a matter as speculative as the nature of the mind was. Also, Hume made a central idea flourishing — the idea that mental complexity could be, through observation, reduced to a set of inter-associated rules. His work would be of the utmost importance for inspiring the forthcoming philosophical Kantian doctrine, which came to establish throughout the nineteenth century and from which current approaches to AI are undeniably rooted.

In the nineteenth century, departing from an incompatible perspective between logic and experience, the emerging doctrine (was) grounded (on) the idea that logic and experience were facets of one another. Following Hume's hypotheses, there would be a set of innate biological rules, which would apply to extract meaning from experience. Hence, experience could only be object of understanding due to these rules, and, reversely, rules were only useful due to the existence of an experiential content. Consequently, the focus of attention shifted from incompatibility to the search of the "geography" of the overlapping area between content (experience) and structure (rules).

In the second half of the nineteenth century, this content-structure interplay became object of research under different approaches and methodologies, which sought to be reportable and replicable. Phenomenology, founded by Husserl in northern Europe (Husserl, 1999), aimed at providing a logical and coherent framework for extracting meaning from the unimaginable large, supposedly infinite, set of possible experiential configurations of human activity. The "phenomenon" was precisely the result of structurally reducing subjective experience to a common meaningful ground, or nucleus. The pleonasm "subjective experience" is here employed to highlight the singular dynamics of experience, where it is not only *what* is experienced that matters, but also *how* it is experienced by the subject.

> The challenge of finding a structure in which an infinite set of inputs would meaningfully fit in was as early driving scientific curiosity. Still it is today. It is precisely the nature of this interplay that lies at the core of AI, where experience is denoted by the inputs that perturb the system and structure is denoted by the stable operations that constitute the algorithm.

In order to glimpse upon the *how*, or the structure, phenomenology largely relied on the power of the narrative — the means by which the subject would express the meaning of actions. In this period, meaning was seen as the underlying unitary force of mental and behavioral movement. Thus, if the narrative comprises the meaning of experience, if the narrative uses language to express the meaning of experience, and if thought was considered to *be* this inner abstract representation of experiential reality; then, language was assumed to be the tool of thought. By the end of the nineteenth century, this idea unfolded into several research and philosophical branches, namely Philosophy of Language and European Psychology.

Through a detailed and logical analysis of language, Philosophy of Language specialized in decomposing the meaning-and-structure interplay of speech as a means to glimpse into how it [meaning-and-structure interplay] leads to the formation (and understanding) of narratives and *thought* itself (Morris, 2006). Philosophy of Language, thus, accompanied the ideas of a mechanical decomposition of processes and functions, having had a profound impact on how to reproduce the innate act of meaning-making and meaning-understanding. As of today, the decomposition and re-composition of language, under the principle of compositionality, is still embraced not only to foster the understanding of language and communication, but also to reproduce its structural and developmental properties within the scope of artificial systems (De Beule and Bergen, 2006). European Psychology, also relying upon the narrative, attempted to combine a structured model of behavioral development with the particular experiences of each subject. The idea was to understand the *self*. With slight variations across authors, the *self* was generally conceived as an integrated and nuclear abstract entity underlying the dynamics of the mind and as the carrier of the *phenomenon*. The mind could only be grasped if experience was abstracted into clusters of meaning, until a nuclear common ground was reached that explained behavior in all its observable facets. Initiated by Eugen Bleuler and Sigmund Freud, it was followed and updated by posterior generations, where Jung was a prominent figure (Bair, 2004).

Simultaneously, but bearing an opposite perspective over the methodology to be used, American Psychology (behaviorism), was materializing the idea that the complexity of mind should be disregarded in favor of an observable behavioral complexity. Observable behavior was already then being proved to be decomposable into an ordered sequencing of stimuli, responses and reinforcement rates (Commons and Liu, 2017). Behaviorists assumed that the essence of the mind should not be in question, as statements about the mind were equivalent to statements about behavior and mental states were equivalent to dispositions to behave (Nath, 2013). The behaviorist movement was early initiated in the beginning of the nineteenth century with the work of Edward Thorndike, but only slightly later coined by John Watson in the first half of the nineteenth century. Relevant work was later conducted by Ivan Pavlov and B. F. Skinner. Behaviorism provided that behavior was measured (and reproducible) and that relevant factors for modulating responses were uncovered (Nath, 2013).

Jung's biographical information (Bair, 2004) suggests that his work might have constituted a link between the analytic approaches dominating in Europe and the behaviorist approach dominating in North America. Contemporaneous with Thorndike, Jung made possible that the qualitative analysis of mental phenomena carried by the *self* was measurable and that its understanding was reproducible. Furthermore, his method was valid when applied to humans, whereas behaviorism had only been so far validated with non-humans. Jung devised an empirical method based on stimulus-response paradigm, measuring reaction times and galvanic skin response during analytic sessions. He was able to show that the body is an inherent part of the *self*, containing its physical, observable, tangible dimension. The connection between European and American psychological approaches was further impelled and highlighted by the first experiments with neurophysiology, which took place later in the nineteenth century. The first conclusions about the modulation of brain waves in response to the environment, in non-humans as well, were extracted by the Polish scientist Adolf Beck in 1890, paired with the first recordings of cerebral activity were published by the psychiatrist Hans Berger (Coenen and Zayachkivska, 2013).

Berger's work constituted the starting point of a scientifically and clinically fruitful era. It was advocated that the observable and quantifiable set of physical properties displayed by the brain reliably represented the dynamics of the *mind*, which it could substitute for research and analytical purposes (Tudor et al., 2005).

In retrospective, two major ideas were revealed. The interplay between content and structure lies at the core of the complexity of the mind, and the complexity of the mind could be grasped by decomposition of brain processes and functions.

In the middle of the twentieth century, it resulted that the correlation between mind and brain was getting progressively stronger and consecutively corroborated. The multidisciplinary mechanical view underlying mental processes was starting to produce endless debates about whether the mind could be, or could not, and to what extent, reproduced in a machine. It was in such an environment imbedded within a canvas composed of measurable variables and outcomes that early approaches to AI appeared.

1.1. Basilar concepts involved in Artificial Intelligence

Independently of the conceptualization of the brain, the object of AI was soon and generally accepted to be natural intelligence, and intelligence could be broadly defined as the ability to adapt and succeed. It implies a constant interaction between the organism (structure) and its environment (content), where experience is the exchanging card. Furthermore, intelligence follows the unidirectional flow of increasing the complexity of adaptive behaviors, requiring that there are changes in the organism and in the way the organism deals with the environment, as experience proceeds. Successful changes are adopted and maintained and set the reference for posterior changes (Dawson-Tunik et al., 2005; Elman, 1993; Inhelder and Piaget, 1958).

The first attempt to implement a system resembling how adaptation is processed was conducted by the English neurologist Ross Ashby, who developed the Homeostat, in 1948. His theoretical framework “brought together physical, biological, and psychological theory in a novel and powerful form, one that he would credit Artur Rosenblueth, Norbert Wiener, and Julian Bigelow, and G. Sommerhoff for having independently discovered it in their own work” (Asaro, 2008). Ross Ashby was the first to apply mechanical concepts, such as equilibrium and amplification, to understand and implement the basic dynamics of adaptation. The author’s major claim was that this mechanistic view would justify and deal with the simplest and the most complex forms of behavior produced by an organism within a single explanatory framework.

He elaborated on the “simplest nature of adaptation, as a route from simple physiology to complex learning”, consistently with the decomposition method early adopted.

Ashby undertook an epistemological approach by breaking down psychological and mental processes as essentially physical and chemical ones. The simplest nature of adaptation would comprise flexibility (search for equilibrium in the face of perturbations) and stability (attaining a new representation of equilibrium) and was operationalized equivalently to Sommerhoff’s directive correlation and Rosenblueth, Wiener, and Bigelow’s conception of negative feedback (Asaro, 2008).

Nowadays, machine learning embraces a similar goal: that of ascribing an optimal compromise between flexibility and stability of a system that is to be perturbed by a determined environmental set. General AI denominates the advent of taking machine learning to the extreme of the perturbations being any, because the system itself is ideally robust enough to (successfully) account them *all*. This observation was early driving scientific curiosity, in the beginning of the nineteenth century. The solution ideally implies that both flexibility and stability are maximized. Another important idea is complexity. Machine learning looks forward to solve highly complex real-world problems. General AI has the goal of solving problems yet unsolved by humans, or of solving problems better than humans.

> Nonetheless, the *route from simple to complex learning* is different from and is more than *learning complex problems*. Solving complex problems implies that learning has already matured to a certain point of complexity, whereas the route from simple to complex learning implies that learning gets more complex as experience proceeds (adaptation).

Given the above, the concepts of flexibility, stability, learning, development and complexity are taken as basilar for the simulation of intelligence and adaptation. But how are they defined and how can they be defined into a coherent and single framework?

1.1.1. Flexibility and Stability

Generally speaking, flexibility is defined by the Oxford dictionary as “the ability to be easily modified” or “willingness to change” (“Flexibility,” 2010). From a psychological perspective, it is a slippery construct to define. It can be attributed to the capacity of a person to adapt to situational demands, to shift among different perspectives of the same situation, to reconfigure mental resources, and to balance competing needs and desires, all including environmental features in coordination with personal features (Kashdan and Rottenberg, 2010). In terms of personal features, psychological flexibility has been attributed, as well, to a multitude of concepts and constructs, namely ego-resiliency, response modulation, self-regulation, and executive control. The latter [executive control] mostly lies within a cognitive perspective of flexibility. Cognitive flexibility basically names the ability to switch among tasks and/or perspectives with ease (short time intervals) and success (low error) (Scott, 1962). Executive control reflects the integrated functioning of a wide cortical frontal area that is the result of several cognitive capacities, such as selective and sustained attention, working memory, and recall (Kashdan and Rottenberg, 2010). Neuroscience data further reveals that cognitive flexibility is correlated with variability in brain processes, which is innate and associated with a healthy cognition (Armbruster-Genc¸et al., 2016; Kashdan and Rottenberg, 2010).

Considering an engineering scope, flexibility is a composite variable that results from time, effort, cost, and performance of the system, and different from adaptability, agility and changeability (Magalhães, 2014). Flexibility has been considered as an inherent operating property of a system, which is capable of altering its internal configuration to respond to a new situation. Adaptability is the capacity of the system to effect internal changes so as to deliver its intended functionality (Schulz et al., 2000). Agility is the capacity of the system to implement changes rapidly. Changeability is a characteristic linearly associated with the set of possible change paths the system has. Two types of change — flexible and adaptable changes — which are not mutually exclusive and are based on the agent of change have been identified (Ross et al., 2008). Flexible changes are produced by external triggers, whereas adaptable changes are produced by internal triggers. Changeability is related to measure flexibility based on the degrees of freedom the system has.

At the level of analyzing the change of a system, the principle of change propagation assumes that change in one element causes change in its closely related elements, modifying the network. Unless the elements that have been changed can be clearly identified, flexibility becomes a latent construct, meaning it cannot be directly observable or measurable. This is congruent with (neuro)cognitive findings. Cognitive flexibility is difficult to operationalize since the behavioral predictors and neural correlates underlying it are not clearly identified. Research concerning cognitive load theory points towards an intricate interaction between external and internal learning, behavioral, and experiential factors (Duran et al., 2018), and research in cognitive neuroscience points towards a predisposition to change of an entire brain network (Leber et al., 2008). This is congruent with the conceptualization of the brain as an entire network of internal dependencies, as formalized by complex system's theory (Mitchell, 1998; Smith, 2005).

In computational terms, flexibility is the characteristic of a "device or program that can be used for multiple purposes, rather than a single function" (Christensson, 2014). In an artificial learning system, which is built to solve problems automatically, the first step is to fine-tune the model according to the (particular) set of problems it is intended to solve. Afterwards, the system is left with the task of solving them. Flexibility in an artificial learning system is usually related with one of two abilities. It is seen as the ability to generalize well to yet unseen cases, also called generalization ability, and/or the ability to solve an increased set of problems, switching between problem-solving functions as different problems are presented. The latter ability is closer to the definition of cognitive flexibility (task-switching) and has ever been way more problematic to implement than the first (French, 1999). In terms of the generalization ability, flexibility of an artificial learning system is associated with a specific performance measure: the variance of the model. The variance is the ability of the model to capture many features in the dataset, eventually leading to overfitting and failing to generalize well (Geman et al., 1992). Consequently, a system that is more flexible and less error-prone in the training set, in the sense of a higher variance, will only solve problems that are very similar to those it has been trained for, and will only solve a particular domain-specific set of problems.

Nonetheless, the variance is mostly associated with the generalization ability of the model for a given task, not so much associated with the ability of the model to learn multiple tasks and to switch among them with ease.

Learning multiple tasks requires the model to learn multiple features of the dataset in a segregated way, by attributing a subset of those features to particular tasks. Switching among tasks requires that flexibility is coordinated with stability.

Stability is defined by the Oxford dictionary as a state of not being liable to undergo any physical change, or firmly established ("Stability," 2010). The same principle holds for every domain of knowledge. Stability is the ability to keep something constant and unchanged. From a cognitive perspective, stability is defined as the ability to minimize the influence of distractors, allowing the subject to focus on a determined task (Armbruster-Genc¸ et al., 2016). In engineering terms, synonyms of stability are, for instance, robustness and uniformity (Magalhães, 2014). Robustness characterizes the portions of the system that are not affected by a changing environment, whereas uniformity means that the system achieves a similar performance within a determined flexibility range. In an artificial intelligence system, stability is the ability of the system not only to form unchanging representations, but to keep those representations unchanged even if in the presence of disrupting inputs. This corresponds to a low variance, which means that the relevant features have been captured, while irrelevant features are discarded. Hence, stability (low variance) can be seen as the opposite of flexibility (high variance). Another measure usually associated with the stability of the system is the bias. A high bias indicates that the model captured well, or is stable in representing a few features in the dataset, but failed to represent other equally important features (Geman et al., 1992). It represents more the error of the model rather than its stability. The bias/variance dilemma is a very well-known problem in machine learning. Models can be decomposed into these two components. The idea is to minimize both — a low bias indicates that the model captured all correct and relevant features in the dataset; a low variance indicates that it did not over-fit.

Yet, as mentioned, the compromise between bias and variance in a machine learning model comprises its generalization ability, not its ability to switch among tasks with ease.

In the natural learning system, flexibility and stability are maximized. Yet, as processes are not observable from the outside but exist in a self-contained structure, they are very difficult to reproduce in an artificial system. At first, flexibility and stability appear to be dynamical and, more importantly, complementary processes. The more flexibility, the less stability, and vice-versa. Hence, if they are seen as opposite sides of the same coin, what is at stake when maximizing flexibility and stability of a single system?

> Maximized flexibility and stability refer to the capacity of the natural system to form many "low bias and low variance" representations of the environment.

Then, is the maximization of flexibility and stability achieved at the level of the system's structure of at the level of its functioning? Maximizing flexibility and stability at the level of the structure implies building a system's structure that is 50% stable and 50% flexible, which says very little to nothing about the dynamics of the process.

> Maximizing flexibility and stability at the level of the system's functioning implies identifying 100% of the moments when stability should overweigh flexibility, and otherwise.

Cognitive neuroscience research, although far from a coherent and clear understanding of the coordination between flexibility and stability, has provided some recent evidence that helps discriminating these two processes in the brain. Flexibility and stability seem no longer to be heads and tales of a single process, but to hold some sort of relative independency. For instance, brain regions like the basal ganglia, nucleous accumbens, prefrontal cortex and posterior parietal cortex resulted as independent sources of prediction of flexible behavior (Leber et al., 2008). Furthermore, mean brain activity (associated with stability) and variability in brain activity (associated with flexibility) are, as well, essentially independent factors, which suggests a likely essential independence of stable versus flexible processes. What further seems to happen is that either one or the other portion of the network seem to be strengthened at the expense of the other (Armbruster-Genc¸ et al., 2016). Hence, subjects who perform better on task switching, showing decreased error rates, perform poorer on distraction inhibition, showing increased reaction times, and otherwise. Network portions are associated, but independent.

1.1.2. Learning and Development

According to cognitive and psychological perspectives, learning is the process through which an organism acquires knowledge, by forming representations. As stimuli are repeatedly exposed, the organism updates the representation of that stimulus, until a reliable and stable representation is established. Each representation is acquired by a substitutive and continuous process, depending on a number of external and internal factors (Commons and Liu, 2017; Duran et al., 2018). More adaptive representations are gradually formed and gradually substitute less adaptive ones.

> Given a certain situation, an organism is able to form a certain number of representations, creating different categories (Damon et al., 2016; Dueker and Needham, 2005; Sloutsky, 2010). Another natural capability is that of forming more or less complex representations, producing wider or narrower representations of the environment (Commons and Pekker, 2008; Sloutsky, 2010).

Development is the process through which an organism forms increasingly complex representations of a given situation, or problem. Differently from learning, development has been observed to progress throughout qualitative and discrete spurts, not continuously (Case, 1987; Commons and Pekker, 2008; Dawson-Tunik et al., 2005; Inhelder and Piaget, 1958). It is associated with the existence of stages of performance, or plateaus. Each stage is assumed to rely on a functional and behavioral pattern. Each pattern, in turn, contains the rules and operations that are particular and descriptive of that stage. As an organism attains a certain developmental stage through the course of its maturation, it will most likely perceive and act upon its environment according to the internal dynamics of the higher stage attained, which is the most adaptive. Hence, each stage acts as an attractor of the system, conferring it stability in the way situations are perceived and acted upon.

Increasing in stage of development occurs through a process called stage transition. While stages of performance are characterized by performance invariance, a transition from one stage to the next is characterized by an unstable pattern that bounces back and forth between two adjacent stages (Dawson-Tunik et al., 2005). It implies that new functional and behavioral patterns are formed throughout an organism's developing lifetime, or maturation. New patterns are assumed to *emerge*, although the specifics of stage emergence are very difficult to trace and reproduce. According to a complex system's perspective, emergent stages, or patterns, result from a process of self-organization. Self-organization is the ability of the system to update and re-organize itself, according to the perturbations it is exposed to (Mitchell, 1998; Smith, 2005). Here lies the second major difference between learning and development. While learning substitutes less adaptive representations of a situation, the transition for higher stages do not substitute or eliminate lower-stages. The organism remains the possibility of reactivating lower-stages, whenever the environment so requires (Fischer, 2008).

1.1.3. Flexibility and Complexity

How are flexibility and stability associated with learning and development?

> Interestingly, in natural systems, there appears to be an underlying universal, structural, biological factor that explains and discriminates flexibility and stability of acquired representations, as learning and development proceed.

For instance, at the same stage, as individuals learn, there is a clear universal tendency to form certain categories of representations over others, as the individual interacts with the environment (Banks and Ginsburg, 1985; Farroni et al., 2005). Also, evidence shows that the complexity of representations progresses according to a specified developmental universal sequence (Commons and Pekker, 2008; Sloutsky, 2010). More complex representations go hand-in-hand with storing more information about a situation, thus, the situation is handled with higher accuracy and higher adaptive potential (Sweller, 2004). More problems and more complex problems can be solved at higher stages, expanding the range of opportunities for success and adaptation (Leite, 2016). Hence, increasing in stage allows for the system to become increasingly more flexible in the way it deals with the environment, while preserving the stability that characterizes performance within stage (Dawson-Tunik et al., 2005). Indeed, cognitive complexity and cognitive flexibility seem to be two fundamental and associated properties of cognition. Whereas cognitive complexity is a structural property, cognitive flexibility is a dynamical property. Cognitively simple subjects tend not to flexibly reorganize information, whereas cognitively complex subjects are more amenable to information reorganization, with gains in how they yield new attributes and concepts (Scott, 1962; Sloutsky, 2010).

Behavioral and neuroscience data actually suggest that flexibility and stability need to be associated with something else other than learning to perform (a set of) tasks.

On the one hand, evidence suggests that switching among tasks, or cognitive flexibility, comes at the expense of increasing the cost of performance, both in terms of time and accuracy. On the other hand, although increased flexibility seems to be detrimental for accurate performance, there is evidence that flexibility increases from childhood to adulthood, demonstrating its adaptive potential (Armbruster-Genc¸ et al., 2016), and that flexibility is a protective factor from psychopathological conditions (Kashdan and Rottenberg, 2010). Furthermore, greater variability levels of brain functioning, associated with cognitive flexibility and task switching, are linked to better performance specially in complex tasks (Armbruster-Genc¸et al., 2016).

This might explain why increased cognitive flexibility is detrimental in some conditions (namely for performing a set of tasks of the same, low complexity) and favorable in others (namely for performing tasks that increase in complexity). It all might depend upon the a priori complexity of the task and on the stage of development of the individual in question (Duran et al., 2018). In artificial learning systems, an increased variance is also associated with the complexity of the task to learn, given that the model is required to capture more non-linear features in the dataset. Nonetheless, in regards to the coordination between flexibility and complexity of reasoning abilities in artificial models, it has been suggested that current approaches in artificial intelligence fail because they fall on a reductionist assumption that does not consider the main characteristic of human brain — to solve simpler tasks before more complex ones, by performing simple actions before more complex ones (Commons, 2008).

1.2. An adopted definition and inter-relation of basilar concepts

Given the above, flexibility and stability, as well as learning and development, can be seen from the perspective of two axes: a horizontal and a hierarchical axis.

Learning is represented horizontally, indicating the formation of several representations within stage. Development is represented hierarchically, indicating the formation of more or less complex representations across stages.

Learning within stage implies 1) flexibly updating each representation as experience proceeds in a continuous and substitutive manner, 2) stabilizing that given representation as it acquires sufficiently descriptive and operative power, and 3) initiating the formation of another representation of the same complexity in case the input set so requires. At the same stage, flexibility and stability are associated with task switching, where the dominating criteria for improving performance lies upon the categories of representations that gain preference over others.

Thus, stability defined irrespectively of stage of development, or horizontally, accounts for fixing diverse representations that might communicate and eventually overlap, but that do not interfere with one another.

In architectural terms, both biologically and computationally, this corresponds to having multiple regions, or modules, operating simultaneously and in parallel, composing one stack of modules (Alnajjar et al., 2012; Bressler and Menon, 2010; Mengistu et al., 2016).

Differently, *learning* in a *developmental* way consists of creating successive and hierarchically more complex representations of the problem space within boundaries of complexity defined by stage. Hierarchical flexibility accounts for reorganizing information in a way that more complex representations of the environment are formed.

> Stability defined in a stage-like manner, or hierarchically, consists in setting the complexity of the formed representations and assuring that more complex representations of a given problem do not interfere with less complex representations of the same problem.

Less complex representations are protected from "upwards" interference and remain available throughout the course of maturation of the system. In architectural terms, this corresponds to having hierarchical stacks, where each higher-order stack emerges from the lower-order stack (Commons, 2008). Stacks are the computational synonym of stages, which are seen at behavioral (Case, 1987; Commons and Pekker, 2008; Dawson-Tunik et al., 2005; Inhelder and Piaget, 1958) and biological levels (Mengistu et al., 2016; Taylor et al., 2015). In computational terms, hierarchical stacks are beneficial for the cost of the system (Elman, 1993; Mengistu et al., 2016; Norris, 1990).

2. Connectionist Models

Connectionist models, or Artificial Neural Network models (ANN), are artificial learning models architecturally similar to the basic arrangement of a natural neural network. They are composed of computational units and connections linking those units. ANN are a strong and widely used class of artificial learning systems. They were created in the second half of the twentieth century and were soon revealed to be a major breakthrough in the pursuit of AI. A retrograde perspective shows that this breakthrough lies upon the fact that neural networks transformed the space of thresholds and rules (as in the homeostat or in cellular automata) into a composite continuous function. It was differentiability, the property of continuous functions upon which learning in ANN was devised, that conferred on these models a great advantage in terms of flexibly adapting to subtle perturbations. ANN models were initially experimented in its simplest form, called the Perceptron.

2.1. Perceptron

The mathematical formalization of a perceptron was proposed by McCulloch and Pitts in 1943 and implemented by Rosenblatt during the late 1950's (Rojas, 1996). The Perceptron was created to computationally represent a neuron. It is composed by a computational unit that receives information from the inputs [x_1, x_2, …, x_n] and generates transformed information as the output [y]. Connections that link inputs to the computational unit are attributed a real-valued scalar, called weight. In order to simulate the synaptic strength, connection weights can be of either sign, corresponding to excitatory or inhibitory synapses. Once inputs reach the computational unit, two different operations are performed. First, the weighted information conveyed by each connection is (usually) summed. Second, this weighted sum is filtered by a function, called the activation function. In a single perceptron, this activation function was initially set to be a discontinuous step function, and later substituted by a continuous function, namely the sigmoid function. Nowadays, the only requirement of the activation function is that it is continuous and differentiable. Another weight, called bias, is associated with the computational unit, introducing a shift in the y-axis of the function. It is considered another input and attributed the value 1 (Figure 1).

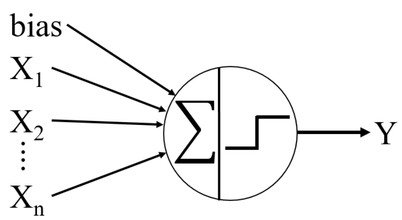

Figure 1 —Representation of a Perceptron with the step function as the activation function

Soon after the development of the Perceptron model, it was proved that the inclusion of more than one unit was computationally possible and efficient for solving more complex problems, creating Artificial Neural Networks (ANN). A Perceptron network thus became a sequence of horizontally positioned activation functions, all receiving input information and transmitting it to the output Figure 2. Each unit configures a specific characteristic of the input set. This layer of computations is called the internal-layer, as it is located in between the input and the output layer. In these cases, biases were linked to each computational unit.

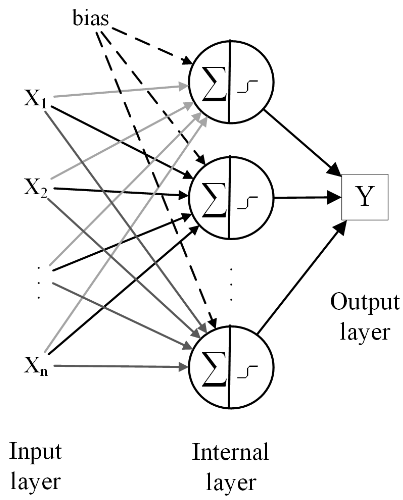

Figure 2 — Representation of a Perceptron Network with more than one computational unit, with a sigmoid activation function

2.2. Multi-Layer Networks

A Multi-Layer network (MLN), or hidden-layer network, is simply an ANN composed of more than one internal-layer of computations. With hidden-layer networks, the implementation of hierarchical processing in connectionism began. With the growth in hierarchical structural complexity of ANN architectures, any non-linear function was proved to be accurately approximated. Nowadays, classical ANN's are viewed as a cascade of activation functions that transform the input space into another space that is the problem solution. This cascade of activation functions imbeds the network with a greater capability of distributed representation and results in more powerful learning capabilities. Importantly, connections between units form the composite function and are basilar to the learning capabilities of the model. The number of layers, the number of units per layer, and how units are connected form the network structure, topology, or architecture. Units can be all connected in a feedforward manner, from the input to output layer (Figure 3), only some units connected to each other, and/or having recurrent connections, which means that a unit is connected both to itself and to other units. Architectural issues assume a great importance in ANN, subtitling the denomination of this computational paradigm — connectionism.

Yet, what seemed to be a good solution for optimizing neural networks with one or two hidden layers became inappropriate when dealing with networks with more layers of computation, necessary for solving more complex problems, mapped by highly varying functions (Erhan et al., 2010). The increase of the number of layers led to poor generalization for representing some functions (Bengio, 2009), leading modelers to limit architectures to one or two hidden layers (Arnold et al., 2011), called shallow architectures. Shallow architectures are, then, assumed to be suitable only for learning relatively complex problems.

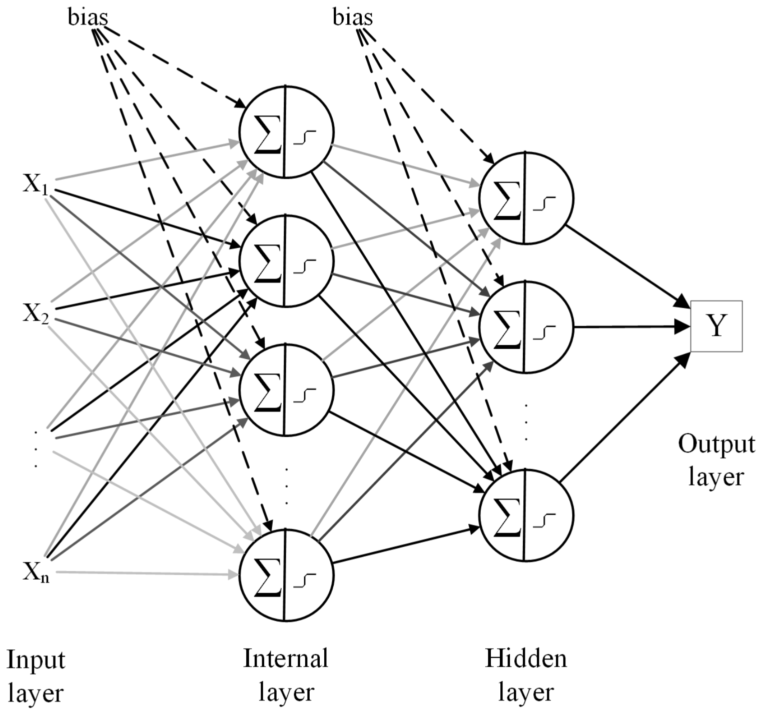

Figure 3 — Representation of a Multi-Layer Feed-Forward Neural Network

2.3. Modular Neural Networks

Besides the problems in learning procedures faced with MLN, another downside was that many features of connectionist models could now be experimented and combined, resulting in an absurd number of possibilities for building a network for a given problem. Modular Neural Networks were employed to reduce the risk of a poor solution. Modular Neural Networks are composite structures of neural network models, in which each neural network is a module of the global system and works as an elementary unit (Figure 4). Also called Stacked Generalization procedures, this option is commonly used to overcome the limitations of individual component networks when there is insufficient training data, when the training data carries a lot of noise or when it is highly expected that the learning algorithm will not find the optimal solution (Dasaratha, 1996; Ting and Witten, 1999). Specifically, MNN were initially used as ensemble techniques (Zhou et al., 2002), feature extraction techniques (Wang et al., 1998), and multi-class classifiers (Anand et al., 1995). Moreover, the fact that different networks perform differently on different regions of the input space leads one to reasonably expect that a combination of models might be more suitable to learn a complex problem.

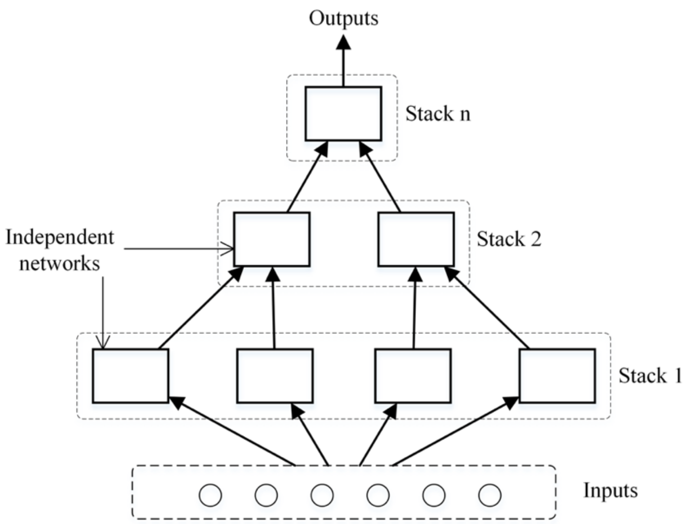

Figure 4 — Representation of a Modular Neural Network

2.4. Generative Architectures

There are two types of generative architectures in neural networks. One type is associated with unsupervised problem-solving. It consists in creating a model that replicates the dataset such that it learns to represent its relevant features. In supervised learning, generative architectures are those neural network structures which topology is not fixed. Generative architectures of this kind allow for the structure of the system to depart from a minimal complexity structure that evolves as more complex problems are presented to it, so as to improve its performance. A known generative architecture of the second type (of interest to the present dissertation) is called cascade-correlation (Fahlman and Lebiere, 1990). Hidden units or pools of units are added sequentially. They are trained in parallel, not interfering with the active network. Whenever there is no more progress on training of these candidates, the one with best correlation with the output is selected and becomes part of the active network.

Generative architectures of this kind were developed to overcome some of the currently known limitations of neural network models, namely in what refers to the best topology. Namely, a cascade-correlation algorithm dispenses with guessing the size, depth, and connectivity pattern of the network in advance. Also, it builds deeper networks without slowing down the training time and improving accuracy (Fahlman and Lebiere, 1990; Shultz et al., 1994).

2.5. Deep Architectures

Deep architectures, the latest class of connectionist models, has been theoretically proved to achieve maximum flexibility and maximum stability in highly complex scenarios of perturbations. These architectures consist of multiple levels, or stacks, of distributed representations. In 2006, the solution to effective training strategies for deep architectures was found, using algorithms for training deep belief networks and stacked auto-encoders (Erhan et al., 2010; Hinton et al., 2006; Hinton and Salakhutdinov, 2006). It was proposed that using unsupervised learning "could be a way to naturally decompose the problem into sub-problems associated with different levels of abstraction" (Bengio, 2009). Each next stack is trained to encode the original problem in fewer features, where the inputs and targets of each stack are the outputs of the previous stack. This approach can be summarized as an unsupervised greedy layer-wise feature extraction followed by supervised fine-tuning, where the features of the input are learned with progressive degree of abstraction, moving the parameters of the network into the right direction and facilitating learning afterwards (Figure 5).

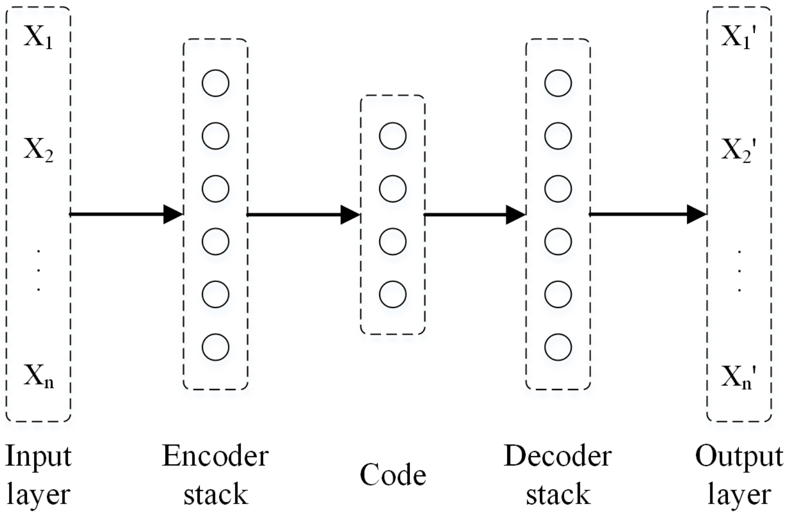

Figure 5 — Representation of a Stacked-auto-encoder

2.6. Simulating Cognitive Information Processing in a Connectionist Model

Learning in ANN consists of updating the connections weights that link units, including the biases, such that inputs are transformed in the desired outputs. This update of the weights during learning is possible due to the existence of a cost function, or error function, which indicates whether the transformation input-output is matching the desired one. The learning algorithm defines how the output error will impact the progressive change in connection weights, until it [the error] is decreased to its minimum, by applying a gradient descent technique. The hyper-parameters that set the learning algorithm and the network's topology are two main factors for successful learning. Both are interdependent and the best choice also depends upon the quantity and quality of input data. These choices eventually allow the understanding of how information processing circuits operate. Thus, they can inform and validate models of cognition under testing.

In fact, cognitive modeling is necessary to understand and validate our understanding of human-level intelligence (Cassimatis, 2012), however, reproducing descriptions does not account for understanding the mechanisms serving behavior (Rijn et al., 2003). It is possible to collect data about human cognition, build fine models that fit the data and accurately predict new observations — it is possible to do all this without actually helping to understand human intelligence (Cassimatis, 2012). This risk is particularly salient when simulating processes with connectionist models. If connectionist models are composed of a sufficient number of inter-connected units and a sufficient number of training cases, they become universal mappers. This means that inputs can always and accurately be transformed into the desired outputs. Hence, there is the risk that a connectionist simulation reproduces the set of collected data, while lacking the capacity to reproduce the set of procedures going on at a biological level. For this reason, *understanding how the human brain embodies a solution to the intelligence problem is (...) in part a cognitive modeling problem* (Cassimatis, 2012), and defining a methodology that allows the drawing of parallelisms between hypothesized natural and resulting artificial systems is absolutely necessary.

As has been briefly described, architectural and algorithmic modifications have been progressively introduced to increase the capability of the model to learn complex functions. However, first, these modifications aimed at overcoming the difficulty of solving highly complex problems, rather than the difficulty of allowing the system to go from simple to complex learning (hierarchical flexibility). Second, for a certain set of complexity problems, flexibility continued to be mostly operationalized as the ability of the system to generalize well to unseen cases, rather than the ability of the system to represent multiple functions and to switch among them with ease (horizontal flexibility).

2.7. Flexibility, Stability, Learning and Development in Connectionist models

In terms of horizontal flexibility, it is important to restate that, across these models, learning capabilities refer to learning within a constrained problem space.

> The capability of networks to learn more than one function remained restricted. The lack of (horizontal) stability concerns precisely the difficulty of the network to protect previously acquired representations in the face of new, dissimilar information.

The difficulty comes from the fact that as new inputs are given to the model during learning, old weights that were important for older information (older representations), are in risk of getting lost; they are not preserved. Forgetting is a problem very early identified in artificial learning models, where connectionist models are not an exception, called stability-plasticity problem. It concerns the duality between the rapid learning about world phenomena (plasticity) and the stability of memory processes (stability) (Carpenter and Grossberg, 1988; McCloskey and Cohen, 1989). The terms catastrophic forgetting or catastrophic interference are the extreme manifestation of what has been previously identified (Goodfellow et al., 2015). Although the cognitive system also exhibits forgetting of many older tasks that do not see a continuity in the subject's experience, this forgetting is not catastrophic and depends upon a variety of factors that go way beyond the mere presentation of a new task (McClelland et al., 1995). This problem has been addressed by several authors without a single conclusion being reached (French, 1999; Goodfellow et al., 2015; Herd et al., 2014; McClelland et al., 1995; Seipone and Bullinaria, 2005).

In terms of hierarchical flexibility, which means learning in a developmental way, it has been somewhat disregarded in connectionist models, especially after the implementation of deep learning approaches. Only a few works were conducted, in which the previous difficulty remains: simpler, less complex representations are in risk of getting lost as increasingly complex representations are formed.

> But why is it important to ascribe developmental abilities to a system instead of building it already complex?

First, development of learning abilities is intrinsically natural. Maturational changes enable conditions which allow learning to be most effective (Dawson-Tunik et al., 2005; Elman, 1993). In terms of simulations with neural networks, it has been shown that a complex task is best learned when the network starts with fewer components (Fahlman and Lebiere, 1990; Seipone and Bullinaria, 2005), which induce severe memory limitations and restricted access to initial inputs (Elman, 1993). It has also been shown that solving a task in a step-wise hierarchical manner allows for decreasing the cost of learning (Mengistu et al., 2016). To date, generative architectures have best simulated learning in a developmental way. These architectures are also able to coordinate the continuous nature of learning with the discontinuous nature of development. However, more complex structures substitute less complex structures. This is the same as saying that more complex stages substitute less complex stages of information processing, impeding the model to reactivate less complex information processing patterns as inputs get, eventually, less complex. It seems that the only way to ascribe hierarchical flexibility to a connectionist model is to find a way to coordinate learning *and* development.

Given the discriminative characteristics of learning and development identified in the present work,

> implementing learning and development in a connectionist model implies that 1) it is built in stacks, where each stack contains a network structure and solves more complex problems than the previous stack, 2) that each stack is built out of the previous stack, in a generative way, and 3) that each lower-order stack is always protected and available, if necessary.

A good example is one concerning the problem of date-calculation. This work began with the study of calculating dates with a distance of months or even years, which begins by calculating dates with distances of days. A three-stack hierarchical algorithm was developed that consisted of the successive application of three simple rules. The second rule derived from the first and the third from the second. It was concluded that the "only way to apply more than one rule would be if rules could somehow be combined into a single, more complex rule, which could be executed in a single step. In a multi-layered net, successive steps of the algorithm could be performed by successive layers of the net". It resulted that all three stages of the network learned relatively quickly, and performance was very good (Norris, 1990). However, this was a rudimentary attempt because lower-order rules were manually protected from the interference of higher-order rules (hierarchical stability), and higher order patterns were not developed on the basis of lower-order patterns. Another good, even better, example was conducted with natural language learning, a few years later. It was found that when the network was allowed to grow in size, changing its structure and incrementally being presented an increased portion of the input set, it was able to learn; otherwise, if the entire input set was given at once to a static network structure, it would fail to learn (Elman, 1993). In this work, the importance of building a developing architecture is highlighted, and one in which higher-order structures are built out of lower order ones. Yet, the formation of increasingly complex representations depended upon the size of the input set. As has been stated in the present work, a universal sequence of development has been identified across domains of knowledge and across individuals of the same species. This means that there is a universal factor triggering the generation of a higher-order structure, which is not (only) the size of the input set.

This factor needs to be clearly identified and implemented as the trigger for stack transition in a developing generative connectionist model, such that the model is not totally dependent upon the inputs it receives. Besides, this must be done in a way that lower-order stacks are protected from the emergence of higher-order stacks.

3. Summary

The present section allowed for extracting some grounding concepts of AI, namely, flexibility, stability, learning, development, and complexity, as well as for coordinating them in a single framework with biological and computational plausibility.

The duets flexibility/stability and learning/development are basilar and intertwined processes of human cognition. The duet learning/development was used to propose different conceptualizations of flexibility and stability, discriminating them as horizontal or hierarchical. Horizontal flexibility and stability are intrinsic to the learning process, in which several representations (functions) are formed as experience with the environment (inputs) proceeds. Learning a representation is a continuous and substitutive process. Learning several representations of dissimilar inputs implies that previously acquired representations are memorized and protected from onwards interference. Differently, hierarchical flexibility and stability are intrinsic to the developmental process and coincide with the formation of increasingly complex representations, as development occurs. Developing from simple to complex representations is a discontinuous and cumulative process. Development implies the existence of stages of performance, or stages of development, and implies that as a new stage is formed, previous stages are not eliminated.

In machine learning, flexibility and stability are seen as complementary processes, measured as the variance and bias of the model. Implementing flexibility and stability in an artificial model, in the sense of forming several stable representations, has been generally problematic due to how representations are formed — in a substitutive manner. The difficulty is always to protect previously acquired representations in the face of new ones. The usual solution for "coordinating" flexibility and stability has been to constrain the problem space that a machine is intended to solve — the stability is manually set, while the system is allowed to flexibly learn within those manually set boundaries. Also, flexibility and stability have been conceived irrespectively of their horizontal or hierarchical nature. A segmentation, clarification, and inter-relation of these concepts, as was here provided, hopefully allows for likely segmenting these processes in an artificial learning model, facilitating their simulation.

In the particular case of connectionist models, structural and algorithmic modifications have been introduced to increase the capability of these models to learn highly complex functions. However, because learning corresponds to the update of a continuous function, the implementation of horizontal and hierarchical flexibility and stability remains problematic. Furthermore, hierarchical flexibility imposes a second difficulty — the coordination of the continuous nature of learning and the discontinuous nature of development.

4. What Next

The "Foundation of a Hierarchical Stacked Neural Networks model for Simulating Cognitive Development" highlights the importance (and the difficulty) of bringing the notion of development to a connectionist model. It is due to the existence of developmental mechanisms that human cognition achieved the complexity it presents today. Development is at the basis of solving complex problems, by forming increasingly complex representations as maturation occurs. Increasing complexity is, thus, at the basis of increasing the adaptation potential. Hence, in order to maximize the similitude between human cognition and artificial models, one should not only tap the flexibility/stability problem in forming many representations, or solving many tasks, but also address the flexibility/stability problem in forming increasingly complex representations. This means to adopt a longitudinal perspective-taking of intelligence and adaptation, simulating the development of cognitive abilities, which is the focus of the present dissertation. The idea is to build an algorithm that not only *learns*, but one that also *develops*. According to the view here defended, horizontal and hierarchical flexibility (and stability) comprise a different operationalization and different, non-exclusive, network architectural approaches. Hierarchical flexibility implies a Stacked Network architecture, where each stack corresponds to a stage of development and operates partially independently from other stacks.

Finally, the idea conveyed in this section about the existence of a structural, invariant, and systematic factor underlying flexibility, stability, learning, and development needs to be clarified and accounted in the design of such an artificial system.

The next step is, thus, to clarify this developmental factor and to understand how the properties of one system (natural) can be transduced into the properties of the other (artificial), particularly in regards to developmental abilities, such that a robust cross-disciplinary bridge is constructed.

Section B

Bridging Developmental Psychology and Computational Cognition: The Importance of a Domain-General Stage Theory

Thanks to a growing trans-disciplinary culture, developmental psychological models have been serving computational models of cognitive development. This provides for artificial learning models to be built based on how humans process information and solve problems, which means developmentally, throughout well-defined stages of performance. At each stage, individuals are observed to solve problems that they were not able to solve at the immediately preceding stage. A more adequate simulation of cognitive development depends partly upon answering a number of questions. The present section responds to some of them and provides ideas for future work to respond to the remaining ones. For instance, which properties of problems interact with individuals' cognitive abilities, such that some problems are solved before others? How should these properties be better transduced into an artificial learning model? What are the properties of artificial learning models that might interfere with the simulated developmental factor? What examples should be borrowed from developmental psychology and which underlying theories should be used in order to delineate possible answers and guide simulation experiments?

This section starts out with the selection and description of a commonly used developmental test for simulation of developmental cognitive abilities — the balance scale test. Next, a review of the behavior assessment rationale (upon which existing simulations are based) is conducted, highlighting some of its strengths and weaknesses. Similarly, the strengths and weaknesses of connectionist simulations is evaluated. It is shown that simulations mimic both the strengths and the weaknesses of the behavior assessment rationale they are based upon, which stresses the importance of questioning the theoretical background.

What are the observed properties of problems that interact with an individual's cognitive abilities? How can they be better integrated with the properties of an artificial learning system?

Finally, an alternative theory is introduced, the Model of Hierarchical Complexity (MHC), the last theoretical update of the assessment method proposed for the balance scale test. It is shown that the MHC covers previous limitations, highlighting what needs to be further clarified. The major contribution of the MHC is that it is a general stage theory that a priori measures the difficulty of problems, ascribing a structural growth to the process of development. It allows for identifying a universal, structural and systematic developmental factor. Clarifications refer to both the developmental cognitive properties underlying problem-solving and the transduction of cognitive properties into algorithmic procedures.

1. The Balance Scale Test

The balance scale test is a developmental test that serves to assess the developmental stage at which children perform, from 5 to 17 years old (Commons et al., 2008; Klahr and Siegler, 1978). This test was primarily developed by Piaget (Inhelder and Piaget, 1958), later revised by Siegler (Klahr and Siegler, 1978) and even later revised by Commons and colleagues (Dawson-Tunik et al., 2010), using the Model of Hierarchical Complexity (Commons and Pekker, 2008). It consists of presenting children a beam with a hinge in the middle, and weights on both sides. By changing the weights and their distance from the center, new problem configurations are created. There are also two supporting blocks on each side, under the arms. Different behavioral assessment methodologies (Commons et al., 1995; Inhelder and Piaget, 1958; Klahr and Siegler, 1978) have been proposed to describe and explain the developmental phases through which children pass as they solve the balance scale test. In the original version of the test (Inhelder and Piaget, 1958), exploratory and verbal behaviors of children were observed, registered, and analyzed by means of an experimenter-biased procedure. Siegler (Klahr and Siegler, 1978) introduced a standardized test procedure and assessment methodology, moving towards an information processing perspective — the Rule Assessment Methodology (RAM).

In Siegler's version of the test, at each new configuration, children are required to predict the state of the scale if the supporting blocks were removed from below. The three possibilities of responses are falling left, balancing, or falling right. Children are observed to go through the following sequence of response patterns: younger children are observed to successfully predict those problems where only weight or distance vary on both sides; later, they are able to solve some conflict problems, where weight and distance vary on both sides, but not all problems; at last, they become able to solve the full range of conflict problems. The last set of conflict problems is called the set of Torque Difference problems, where children must multiply weight by distance on each side to correctly predict the side tipping down. The first set of conflict problems is assumed to be easier for two possible reasons: one is that Torque Differences are greater, which allows for the discrimination of the sides based on perceptive skills (Ferretti and Butterfield, 1986; Jasen and Maas, 2001; Schapiro and McClelland, 2009); the other is that this subset of problems is correctly solved by using the summing operation, instead of multiplication (Dawson-Tunik et al., 2010).

After the RAM has been proposed, the Balance Scale Test received growing attention from AI scientists, who used it to reproduce developmental cognition in artificial learning models. Simulations of the Balance Scale Test have been conducted using symbolic models (Schapiro and McClelland, 2009), and connectionist models (Dandurand and Shultz, 2009; Dawson and Zimmerman, 2003; McClelland, 1995, 1995, 1989; Reyes et al., 1997, 1997; Schapiro and McClelland, 2009; Shultz et al., 1994; Shultz and Schmidt, 1991; Zimmerman, 1999). Symbolic models are based on the coordination of computational rules operationalized by *if-then* statements. A correct coordination of rules leads to the generation of correct outputs. The more rules that are coordinated, the more complex responses the model generates. Connectionist models are based on connecting and quantifying data features that flow from input units to outputs units. They are assumed to capture the continuous nature of learning and gradual increase of performance accuracy (McClelland, 1995, 1989; Zimmerman, 1999) and they also capture developmental spurts, associated with qualitative changes in performance (McClelland, 1995; Shultz et al., 1994). The goal of simulations is to build an artificial learning model that learns to solve the problem as humans do, throughout the same sequence of observed/assessed performance stages.

1.1. Rule Assessment Methodology (RAM)

Siegler's configurations correspond to balance scale problems, which are defined according to their "type" (Siegler and Chen, 2002). In *balance* problems, the same weight is placed at the same distance. In *weight* problems, one side of the scale overweighs the other, with weights placed at the same distance from the fulcrum. In *distance* problems, the same weight is placed at different distances on each side. In *conflict-weight* problems, different weights are placed at different distances, and the side tipping down is that with greater weight, not the one with greater distance. *Conflict-distance* are opposite to *conflict-weight* problems. The side with greater distance tips down. Finally, in *conflict-balance* problems, although weight and distance differ in both sides, the scale balances (Figure 6).

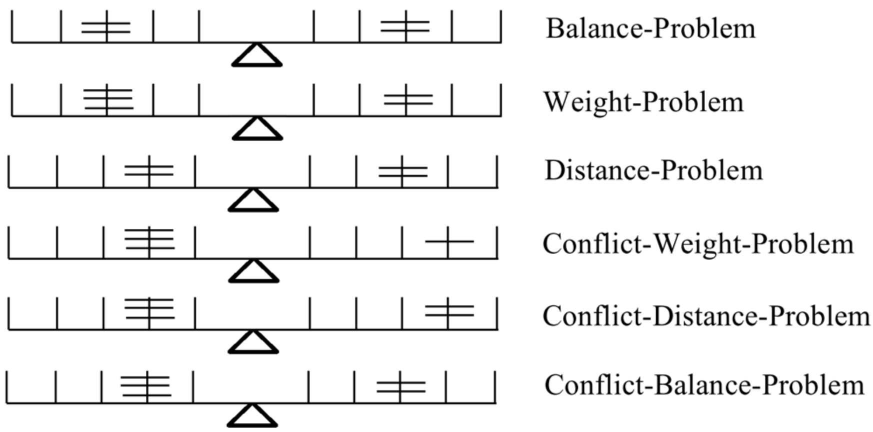

Figure 6 — Example of Siegler's Configurations

Based on these six problem types, Siegler defined a set of four rules (Table 1) that describe problem-solving strategies. These rules are assumed to require increasing cognitive capabilities, having correspondence with domain-specific developmental stages. By classifying human behavior according to one of four rules, Siegler's data showed that human's cognitive abilities progress along discrete stages. For instance, during the second transition, behavior is modeled by a U-shaped learning curve. This means that when children take note of distance information together with weight information, their performance drops to chance level (Dawson-Tunik et al., 2005; Shultz et al., 1994; Zimmerman, 1999). This reflects that children are revising their current knowledge and updating it in order to solve different, more difficult problems.

Table 1 — Siegler's rules / stages

Rule	Mental Strategy	Balance-Scale Problem
I	Considers weight only	Balance problems Weight problems Conflict-weight problems
II	Considers distance as long as weight is the same	Distance problems
III	Considers weight and distance as long as torque difference is large; otherwise, considers weight	Conflict-weight problems Conflict-distance problems
IV	Considers weight and distance and calculates torque difference by multiplying	Conflict-Balance problems

RAM relies on a comparison between the actual and the expected response to each type of balance scale problem, as designated by rules I, II, III, and IV. When 80% of the actual responses correspond to the expected, then, the individual is classified in accordance to that rule (Klahr and Siegler, 1978). The definition of these rules was the fundamental aspect of Siegler's approach. It made RAM a standardized and pioneering method that integrated Piagetian ideas into an information-processing framework (Zimmerman, 1999), providing for a computational basis of human cognition.

1.2. Limitations of the Rule Assessment Methodology

Some limitations are attached to this assessment method. Namely, when the actual responses conform to two rules with the same frequency *or* when the actual responses are less frequent than 80%, the behavior is *unclassifiable* (Siegler and Chen, 2002). In these cases, the identification of the maturation of reasoning abilities is problematic. In every other *successfully classified* case, there is the problem of an arbitrary threshold criterion (80%) (Klahr and Siegler, 1978; Maas et al., 2007; Zimmerman, 1999). This might prove statistically significant in some cases and not significant in others. A statistical method was developed that overcomes this classification problem. It is called Latent Class Analysis (LCA) and uses factor analysis for classifying behavior based on latent variables rather than on overt categorizations (Jasen and Maas, 1997; Maas et al., 2007).

Furthermore, Siegler's assessment methodology has been criticized because it lacks a measure of goodness of fit (Wilkening and Anderson, 1982), criteria for classifying the type of errors, and the consideration of the *torque difference effect* (TD) in classifying performance (Zimmerman, 1999). The TD is a phenomenon linked to information saliency effects (Ferretti and Butterfield, 1986). Larger TD are easier to solve than smaller ones. Other critics refer to test properties and task demands, arguing against the underestimation children's knowledge and the incorrect classification of their mental strategies (Zimmerman, 1999).

Besides these limitations, the present work adds that the major limitation of Siegler's rules relies on the fact that they are categorical, mentalistic, and task-specific, failing to measure the increase in complexity that characterizes the maturation of reasoning abilities. Different configurations of the balance scale are assumed to be more difficult than others because children appear to solve them in different timeframes, measured in years (Klahr and Siegler, 1978).

However, a priori difficulty is not measured. The absence of an a priori difficulty measurement leads to three consequences.

First, although it is argued that rules encompass general aspects that hold across similar problems (Siegler and Chen, 2002), they result from an analysis of human behavior specifically applied to the Balance Scale test. This makes them contaminated by the nature of the problem, task-specific, with limited scope and open to interpretations (Zimmerman, 1999).

Second, rules seem to face some inconsistencies. For instance, *conflict-weight* problems are actually solvable by children who apply rules I, III, and/or IV, as only weight needs to be considered. Also, *non-classifiable* performance is a possibility when two rules are used with the same probability. This points torwards the effect of transition factors in development, which is of the utmost importance, but not covered by the Siegler model.

Third, and the most relevant, it has been observed that a U-shaped curve only occurs in the transition from *distance-problems* to *conflict-weight* problems (from rule II to rule III). Given that a U-shaped curve characterizes *all* developmental jumps (Dawson-Tunik et al., 2005), this observation questions whether there is actually an increase in difficulty in the remaining rule transitions. It also questions whether the properties of problems accounting for a developmental transition have been well defined. For instance, does the transition between rules I and II reflect different stages of development? And what about the transition between rule III to rule IV?

In terms of the properties underlying developmental transitions, further evidence indicates that Siegler's rules do not allow for the discrimination of the factors which explain why problems are solved in different timeframes throughout development. For instance, the TD effect has been assumed to rely on perceptual capabilities (Jasen and Maas, 2001; Schapiro and McClelland, 2009), but is only fully accounted as operative capabilities are matured, such as the use of multiplication. Catastrophe flag theory applied to the balance scale test (Jasen and Maas, 2001) relies predominantly on the perceptive aspects of the problem and on the training the individual has on certain configurations. With training, the individual shows a gradual tendency to attend and integrate different properties of the problem. Yet, at some point, when the use of perceptual abilities is no longer adaptive, the use of multiplication is required.

In sum, the major drawbacks of Siegler's assessment methodology derive from the fact that it categorically classifies actual human performance on the specific tasks of the Balance Scale test, instead of measuring it a priori. Consequently, RAM is framed within a specific problem type, it is context dependent, and cannot be abstracted for measuring human development in other problems/domains. The specificity of such approach will necessarily prove insufficient for simulating the structural dynamics of cognitive developmental phenomena. Second, RAM does not allow for the identification of the factors that underlie cognitive development, neither on the balance scale test nor on any other developmental domain. If the factors underlying developmental transitions are unclear, so are the factors underlying stable stage performance. Both are necessary when a simulation of stages and stage transitions is aimed at.

1.3. Connectionist Simulations of the Balance Scale Behavior Following the Rule Assessment Methodology

Although symbolic and connectionist models have been used for simulating the balance scale test, only connectionist models will be reviewed for two main reasons. First, connectionist models are parallel processors of information by nature, which makes them a fair representation of the parallel structure of information processing composing the brain (McClelland, 1995). The structure of the network can become a great source of understanding of natural networks for information processing. Second, connectionist models can capture the magnitude of expression of a certain feature or input dimension. Inputs are weighted sequentially throughout the network, making these models more suitable for incorporating information saliency effects, such as the TD effect, rather than relying solely on the coordination of rules (Shultz et al., 1994; Zimmerman, 1999). Also, weights are coefficients of a composite continuous function that results from the connections between computational units. Weights are updated by a gradient descent technique, which favors the representation of the continuous nature of learning.

Apart from the current dissertation, existing simulations of the balance scale test are based on RAM. They were initiated by McClelland (McClelland, 1989) under the Parallel Information Processing project, to which the works of McClelland (McClelland, 1995), Shultz and colleagues (Dandurand and Shultz, 2009; Shultz et al., 1995, 1994; Shultz and Schmidt, 1991), and Zimmerman (Dawson and Zimmerman, 2003; Zimmerman, 1999; Zimmerman and Croker, 2014) followed. These works were conducted based on two main manipulations – architectural and environmental. The architectural manipulation refers to the structure of the network. Fully and not-fully feed-forward networks and cascade-correlation networks were tested. The environmental manipulation refers to a training bias that over represented some input patterns, namely weight problems, such that contextual contingencies were simulated. In these experiments, the contextual bias was aimed at replicating the increased experience of children with weight information over distance information. These simulations were driven with three objectives: to simulate 1) rule-like behavior, 2) U-shaped performance on *conflict-weight* problems, and 3) the TD effect.

In the first experiment (McClelland, 1989), the network topology consisted of a not-fully connected feed-forward topology, reflecting the segregation between weight and distance information. McClelland restricted the connections between input and the two internal units, in that weight information went to one unit and distance information went to the other. Weight and distance information was combined only between the internal and the output units. Results showed that the architectural condition turned out to be critical. From the perspective of a coarse-grained analysis, the model captured the expected developmental progression of children. Yet, rule IV was barely attained. Furthermore, McClelland discussed the continuous versus discrete nature of development, following an approach that was supportive of the idea of continuous developmental changes.

In later experiments, McClelland (McClelland, 1995) tested the inclusion of the TD effect (Ferretti and Butterfield, 1986) and questioned the previous assumption of training bias. The test set and learning criteria were modified to accommodate a broader range of possible actions and possibly allow for the TD effect to become pronounced. As a result, although the correspondence between the new model and the proposal of a TD effect was not perfect, it was close enough to suggest a similarity pattern between the model and children's behavior. In terms of eliminating the contextual bias, inputs were represented in a different way. Distance cues were suggested to be more complex than weight cues, in that distance implies establishing a relation between the target (the balance weight) and the fulcrum. Distance was, then, encoded as a relative position among objects: the position of weights on each side and the position of the fulcrum. With this representation, results demonstrated a clear advantage of weight over distance without the need of introducing a training bias. The suggestion for differential information processing was again confirmed, and the complexity of information processing was again taken into consideration. However, rule II performance, which encodes distance, was very unstable.

By the same time, Shultz and colleagues (Shultz et al., 1994; Shultz and Schmidt, 1991) conducted an experiment in which they kept the environmental condition proposed by McClelland (McClelland, 1989), but employed a different topology. They simulated the balance scale phenomena with a cascade correlation (CC) algorithm. The CC is a generative architecture that adds hidden units to progressively suppress the total error. If the environmental bias was introduced, this generative approach was successful in accomplishing all three objectives: 1) stage-like performance, 2) U-shaped learning in conflict problems, and 3) the TD effect. The learning algorithm showed quantitative and gradual changes in performance, demonstrating the continuous nature of development. Simultaneously, the inclusion of hidden units provided for qualitative jumps, showing developmental discontinuities. This architecture was also shown to be less computationally complex and to learn more rapidly (Fahlman and Lebiere, 1990). However, without the environmental condition, the network would not learn rule I and rule II steps. It immediately jumped to rule III, not showing either stage-like behavior or U-shaped learning curve in *conflict-weight* problems.

Furthermore, intuitive networks, identified as those which keep an architecture stable and only learn from examples, perform well on all problems requiring rules I, II, and III, but fail to accurately represent rule IV. The solution encountered that leads the network to use rule IV, while preserving rules I and II, has been to inject a torque function, by using a Function Based Cascade Correlation architecture (Dandurand and Shultz, 2009), or to inject a new assimilation function — weighted *product* instead of weighted *sum* (Reyes et al., 1997). The assimilation function is the first computation performed at the computational units: the function that agglomerates the weighted information destined to that unit. Usually, the assimilation function is weighted sum. The modification from sum to multiplication is basically a shortcut to make the network solving problems based on a new operation. Even if this solution is efficient, none of these works justify how intuitive networks develop to solve problems requiring rule IV.

So far, balance scale simulations were focused on evaluating the *progression* of activation of hidden units, so as to infer how the network develops until it learns to classify all balance scale problems. Differently, Zimmerman's work (Zimmerman, 1999) consisted of a detailed evaluation of the *final* behavior of a fully-connected feed-forward neural network, or of a fully-developed system that already solved *all* balance-scale problems. What mostly differentiates this work from the others is that the focus was placed on the activation patterns of internal units *after* the network has learnt all problems. The goal was to assess the structure of balance scale problems and re-appreciate Siegler's rules. Zimmerman's results show that distance and weight problems lay in the same problem space (hence, the usage of rules I and II), while conflict problems (rules III and IV) lay in a separated problem space. These results question whether weight and distance problems have different structural properties, but fail to show that the TD effect introduces different problem properties to deal with.

Zimmerman's experiments addressed as well the continuity/discontinuity of development, showing that the network solves the tasks via approximating an additive assimilation function. Tasks are characterized by the continuous properties of dimensions rather than by discrete or nominal characteristics. Nonetheless, problems are clearly clustered, which simultaneously suggests that stages are discrete. Finally, Zimmerman argued that, because a fully-connected network learned to solve all problems, the effect of a segregated topology remains undetermined. However, the experimentation with different topologies serves *not* to solve the problems, but to study how a connectionist structure encodes developmental properties in problem-solving. For this reason, what remains unknown is what is missing in existing connectionist approaches such that the simulation of an apparently simple developmental problem has yet been clearly achieved and explained.

1.4. Summary of existing simulations based on the Rule Assessment Methodology

In sum, all models captured a developmental course from rules I to IV, except McClelland's first work, which failed to attain rule IV performance. All captured a U-shaped learning curve in the transition from rule II to rule III, as expected. All captured the TD effect, except McClelland's first work, which was not intended to do that. Therefore, all experiments mimicked the rationale they used to explain performance on the balance scale test and practically all fairly achieved the objectives they aimed at.

Yet, 1) none questioned the absence of a U-shaped learning curve in the remaining transitions, namely from rule I to rule II and from rule III to rule IV; 2) none questioned the properties underlying the TD effect that made it so difficult to simulate and explain with artificial models; and 3) all required the introduction of a contextual bias such that rules I and II were reproduced.

These facts stress the importance of re-examining which properties of problems are observed that interact with an individual's cognitive abilities and how to better integrate these properties with the properties of an artificial learning system.

In fact, a U-shaped curve is observed to characterize *every* stage transition (Dawson-Tunik et al., 2005). As the child encounters a new experience, the conflict between the old and the new needs to be balanced and resolved. This process creates disequilibrium, but constitutes the integration of new information, known as adaptation (Bartolotta and Shulman, 2013). Hence, it would be expected that such a curve would characterize the acquisition and use of every rule, from I to IV, which is not the case in Siegler's observations.

The fact that a manipulation always had to be introduced in the training set so that the models could differentiate weight (rule I) and distance problems (rule II) (McClelland, 1995; Shultz et al., 1994; Shultz and Schmidt, 1991) calls attention to the task-specificity of these rules in contradistinction to the general structure of development. Rule I and Rule II had been distinguished because children solve weight problems before distance problems. However, Zimmerman showed that, structurally, these problems overlap (Zimmerman, 1999). This calls attention to distinguishing between properties of problems that interact with problem-solving abilities and experiential factors. If the goal is to capture the general developmental structure of problem-solving abilities, the details of problem-solving that depend on experience should deliberately not become the object of simulation. The dependency upon the contextual manipulation also indirectly shows that simulations do not yet capture the right developmental properties of problems, as modeled by the RAM.

> Consequently, because all models are fair simulations of the model they follow, it is fair to assume that these limitations should be revised, not by insistently changing parameters of the artificial models, but by revising the rationale that is used to build simulations upon.

In regards to to Rule III and Rule IV, these have been considered to be part of the same problem space — conflict problems — but their difficulty is agreed to depend upon information saliency effects and the effect of training (Schapiro and McClelland, 2009). Oppositely, in existing simulations, rule IV problems have been solved by injecting a different operator (Dandurand and Shultz, 2009; Reyes et al., 1997), not by injecting a different method for evaluating the input set. Hence, no clear boundary between conflict problems and the TD effect has been delineated, neither has a method been proposed for identifying where the boundary can be defined.

Given the above, an argument is laid out that an alternative rationale for assessing human developmental behavior is necessary. We argue in favor of a domain-general, mathematical, objective, non-mentalistic approach that can bridge developmental psychology and computational learning. Domain-specific approaches advocate that growth and development in one domain forge growth and development of adjacent domains or skills (Bartolotta and Shulman, 2013).

Yet, a domain-general approach grants the advantage of assessing development independently of the content, allowing for cross-context comparisons (Dawson et al., 2003).

Cross-context comparisons are not only valuable for coordinating different fields of knowledge, but also for coordinating different methodological fields, as is the case of developmental psychology and computational cognition. A domain-general theory will allow for models of cognitive development to go beyond capturing the particularities of problems, to capture the structural properties that ascribe them different levels of difficulty, as perceived by individuals. Only then, will one be able 1) to identify which factors of problems interact with an individual's cognitive capacity and 2) to identify which properties of the artificial learning models interact with the phenomena we are trying to simulate.

2. The Model of Hierarchical Complexity (MHC)

The Model of Hierarchical Complexity (MHC) is a Post-Piagetian general stage theory of development (Commons and Pekker, 2008). It applies a measurement method — Rasch Analysis — that assesses the performance on tasks differing in hierarchical complexity. The Order of Hierarchical Complexity (OHC) of tasks is a unidimensional measure that predicts developmental cognitive capacity, with correlations above 95% (Giri et al., 2014).

> The major contribution of the MHC was, in fact, the identification of the OHC as a one-dimensional predictor of cognitive capacity (Giri et al., 2014). This measure has been shown to be independent of the environmental circumstances of the testing procedure, as well as independent of the strategies that might be at use for solving problems (Giri et al., 2014).

It only addresses the hierarchical complexity of the task and whether the individual correctly solves it or not. Each OHC has a one-to-one correspondence with stages of development. 17 orders are shown to form a one-way equally-spaced ordinal sequence (Commons et al., 2014b). The estimation of stages produced by the assessment method used in the MHC are clearly demarcated, with robust gaps between stages, and no overlap (Dawson et al., 2003). Recently published evidence points towards the biologically structural underpinnings of the OHC, which will be described with more detail ahead (Chapter II, Section D).

Also, the MHC differentiates between horizontally and hierarchically more complex tasks. Horizontal complexity refers to the cognitive load of a given task, without changing its OHC. For instance, whether an individual sums 3 elements or 12 elements does not change the fact that the individual is summing, but the number of elements of the sum makes the task more demanding in terms of cognitive load (Duran et al., 2018). Differently, hierarchical complexity deals with the coordination of more than one action, or operation. For instance, the distributive law is hierarchically more complex than addition or multiplication alone (Commons and Pekker, 2008). In other words, horizontal complexity refers to how an information set is organized in chain, by modules positioned horizontally, containing invariances. The invariance is due to the same order of hierarchical complexity of problems. Hierarchical complexity refers to how an information set is organized into hierarchical levels of modules. A higher-order level emerges and abstracts the information contained across modules in the immediately lower-order level. Hence, while horizontal complexity might be represented by the number of modules in a level, hierarchical complexity is represented by the number of levels, or orders (Figure 7).

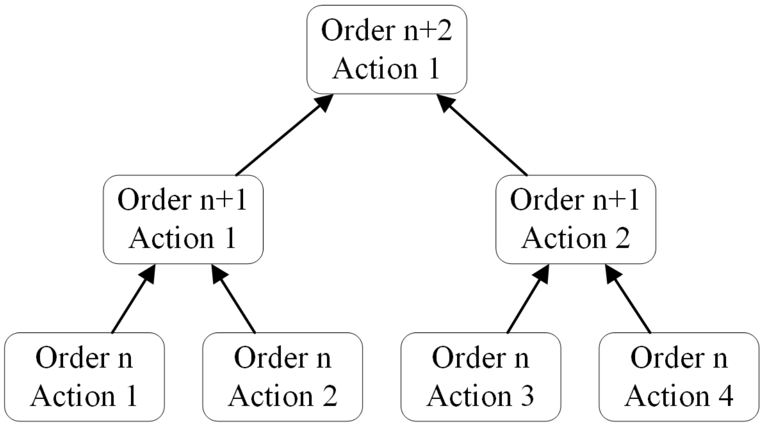

Figure 7 — Non-arbitrary coordination of lower order actions

Each higher-order is characterized by the non-arbitrary coordination of lower-order task-actions, according to the following definitions and axioms:

1. It is defined in terms of two or more lower-order task actions. In mathematical terms, this is the same as a set being formed out of elements. This creates the hierarchy. $A = \{a, b\}$ a, b are "lower" than A and compose set A. $A \neq \{A,...\}$ A set cannot contain itself. This means that higher order tasks cannot be reduced to lower order ones. For example, postformal task actions cannot be reduced to formal ones;
2. It organizes lower order task actions. In mathematics' simplest terms, this is a relation on actions. The relations are order relations $A = (a, b) = \{a, \{b\}\}$ an ordered pair;
3. This organization is non-arbitrary. This means that there is a match between the model designated orders and the real-world orders. This can be written as: Not $P(a,b)$, not all permutations are allowed.

Based on this definition, the hierarchical complexity of tasks is analyzed by reducing the task to its lowest order components. The number of recursions necessary to find these components corresponds to its OHC. The highest order task that individuals can solve, at least once, indicates the

OHC of their reasoning abilities and, consequently, their stage of development (Commons and Pekker, 2008).

2.1. The Model of Hierarchical Complexity applied to the Balance Scale Test

Piagetian theories of human development, including Siegler's, determined that children develop their reasoning abilities throughout four stages, until they attain the fourth and highest stage of formal operations (Inhelder and Piaget, 1958). Hence, the most difficult configuration of Piaget's and Siegler's versions of the test corresponds to a formal stage problem, where the torque effect needs to be solved by multiplying weight and distance on the same side. Differently, the MHC identified a sequence of 17 stages of development (Commons and Pekker, 2008). It showed that only humans attain the formal stage (Commons et al., 2014c), but individual differences indicate that humans can go up until stage 17 (Meta-Cross-Paradigmatic) (Commons and Pekker, 2008). The MHC describes the four Piagetian stages as 11 finer-grained stages, and includes six Postformal stages of development. For instance, differently from Piaget's theory, the MHC discriminates between abstract and formal reasoning. At the abstract stage (MHC's stage 10), children become able to create variables out of a sequence of similar experiences, whereas at the formal stage (MHC's stage 11), they become able to operate with variables. At the forthcoming stages, if they are eventually attained, people are able to create and to operate with systems of variables (systematic stage 12), to create and operate with meta-systems (meta-systematic stage 13), and so on until stage 17 (Commons, 2008). The more fine-grained stages are described, the easier it becomes to reproduce stages of development as levels of information processing in an artificial system. The MHC has updated the balance scale test according to its extended developmental sequence, from the Primary (attained around 5 or 6 years-old) to the Meta-Systematic stage (attained by only a few adults) (Commons et al., 1995). The ordering of problems composing the balance scale test is done according to its measurement system (Commons et al., 2008), and the most difficult configuration of the MHC's version corresponds to a meta-systematic information processing acquisition.

The test consists of a pen-and-pencil multiple choice instrument, in which individuals are asked to complete the information of figures so that the scale balances (Dawson-Tunik et al., 2010). The positions that should be completed are appropriately flagged, so as to remove any exploratory tendency (McFadden et al., 1987). Commons' version also attempts to minimize the influence of the perceptual properties of the stimulus, fostering the use of logical deductions and simple arithmetic for problem solving (Figure 8).

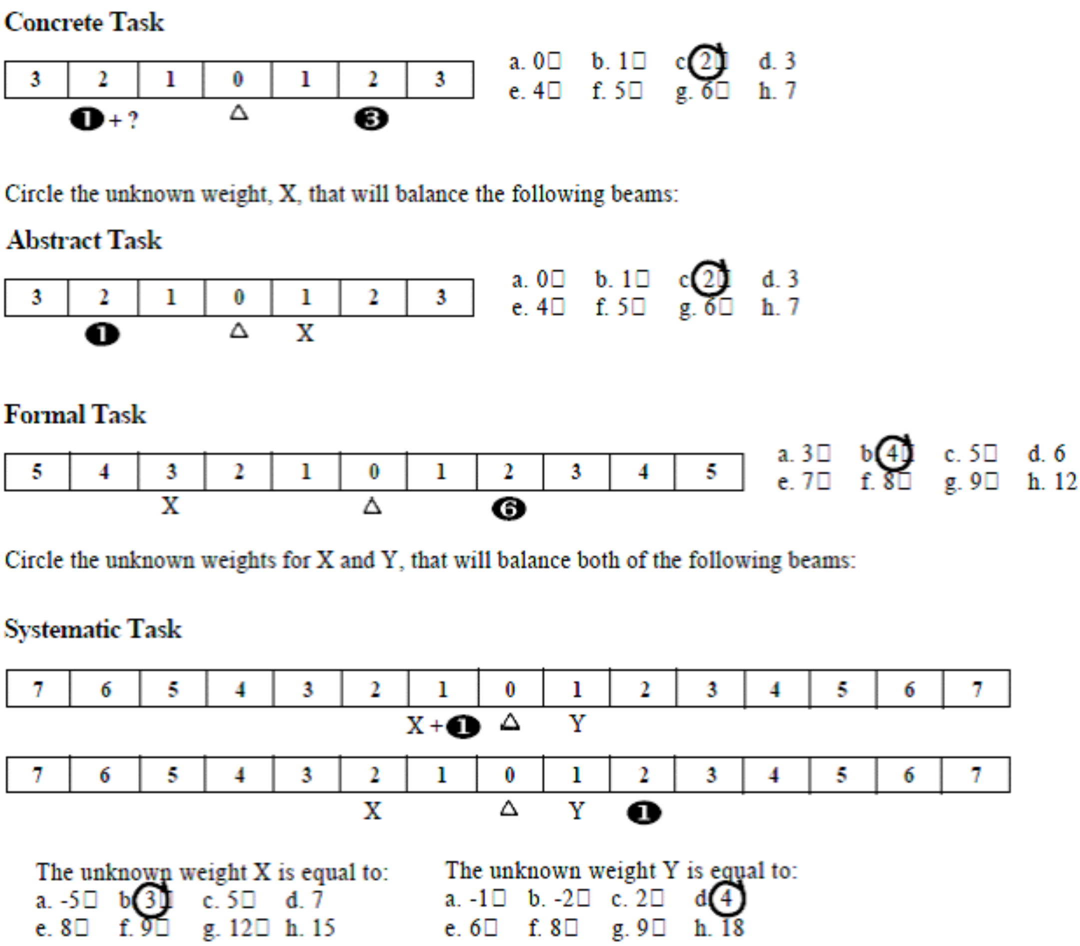

Figure 8 — Commons' version of Balance Scale tasks (Dawson, Goodheart, Draney & Commons, 2010)

Importantly, what most discriminates the MHC assessment method from the previous is that the MHC determines balance scale configurations solely based on the operations necessary to solve each OHC problem (Table 2). It is important to mention that the Primary configurations do not yet require an arithmetic operation, only the determination if number are equal or different on both sides. Meta-systematic configurations present some inconsistencies (not object of the present work), reason why these two configurations have been excluded from behavioral studies.

Table 2 — Operations per OHC problem

Problems	Operation
Concrete	Count how many pegs exist on each side
	Count how many weights exist on each side
Abstract	Sum weight and distance on each side
Formal	Multiply weight by distance on each side
Systematic	Distributive law applied on each side

2.2. The Model of Hierarchical Complexity as an Alternative Theory for Conducting Simulations of the Balance Scale Test

The fact that RAM does not measure the difficulty, or OHC, of tasks has been pointed out as its major limitation, from which other limitations followed. The MHC is assumed to cover these limitations since it is based on a measurement theory, classifying behavior as a function of the measured complexity of tasks and actions. Because the order of complexity of a task-action is exclusive and exhaustively classifies behavior, it establishes a one-to-one correspondence between problems and behavior (Dawson-Tunik et al., 2010), which eliminates *unclassifiable* cases.

Furthermore, the MHC, because it measures a task's hierarchical complexity, shows that rules I and II pertain to same stage, as highlighted by human data and computational data (Zimmerman, 1999). This statement is based on the fact that both these rules only require the manipulation of one variable alone, whether it is weight or distance. Because children are assumed to have more experience with weight information than with distance information (McClelland, 1995), it is expected that they solve weight problems before they solve distance problems. However, it does not change that fact that they can solve problems based on one variable. The most we can assume is that shifting attention to the presence of distance indicates that a transition to integrating both variables will soon occur.

Moreover, the MHC discriminates conflict problems and conflict TD problems, showing that there is a clear transition between them. The model discriminates both stages based on the operation required to solve each subset of problems. Conflict additive problems are those where the sum of weight and distance one each side solves the problem correctly, whereas conflict multiplicative problems — or conflict TD problems — require that the child multiplies weight by distance on each side. Hence, the unclear boundary between the use of perceptual abilities and the use of operative abilities is eliminated. Table 3 summarizes Siegler's and Commons' differences concerning the definition of developmental actions, referencing rules and stages to the problem types identified by Siegler.

Table 3 — Comparison between Siegler's and Commons' assessment methodology

Balance Scale Problem	**Siegler's**	**Commons'**
Balance	I	Primary
Weight	I	Concrete
Distance	II	Concrete
Conflict-weight	I, III or IV	Abstract or Formal
Conflict-distance	III or IV	Abstract or Formal
Conflict-balance	IV	Formal

In sum, because it a priori measures complexity, the MHC:

- Measures a problem's difficulty, which moves away from a mentalistic, task-specific perspective and allows for a one-to-one correspondence between correctly solved problems and stages;

- Explains the absence of a U-shaped curve in the transition between rule I and rule II;
- Eliminates the blurred boundary between conflict problems and conflict TD problems, explaining why a U-shaped curve should exist in the transition between both subsets of problems;

Yet, a few clarifications are needed in further simulations that eventually use the MHC as a theoretical reference. According to the MHC, orders of hierarchical complexity are equally spaced, forming a linear growth (Commons et al., 2014b). This has been interpreted as transitions across stages being equally difficult and requiring an equally deep jump. Yet, the model presents some unclearness as to which factor underlies hierarchical complexity. These factors are simultaneously attributed to:

- Problem dimensions (or inputs): the MHC assumes that the number of problem dimensions grows exponentially with stage, 2^o, where *o* is the order of hierarchical complexity (Commons and Pekker, 2008). This is due to the fact that progressively more elements are combined by means of the coordinating rule R. This implies that the growth in complexity is directly dependent upon the number of problem dimensions that are combined;
- Operations performed: the MHC also states that the order of hierarchical complexity is measured by the number of recursions that the coordinating rule R must perform on a set of primary elements. This growth mechanism points towards the idea that operations conducted from one order to the next are what matter for defining hierarchical orders. Even though, the nature of the coordination rule R is not defined;
- Actions generated (or outputs): in the particular case of the balance scale test, the transition from the formal to the systematic stage relies not only on the number of problem dimensions (number of inputs), but also on the number of solution dimensions (in the systematic problem subset, the individual is asked to find the value of two variables instead of only one value) (Dawson-Tunik et al., 2010).

3. What further is missing in Connectionist Simulations of the Balance Scale Test

If the idea is to build a system that *learns* to solve-problems in a *developmental* way, the concepts of learning and development should be clearly distinguished. This will also allow for distinguishing between horizontal and hierarchical flexibility/stability, respectively. The concepts of difficulty and problem complexity should also be defined, both in terms of developmental psychology and in terms of computational learning.

3.1. Learning and Development

Learning, as a function of experience, concerns the acquisition of adaptive behavioral patterns, when triggering stimuli are presented. Learning can be seen as the optimal linking between patterns of perception and patterns of response. As similar stimuli are repeatedly presented, learning consists in making fewer errors as repetitions occur. Stimulus similarity is another caveat, as it depends not only on the absolute properties of stimuli, but also on how these properties are captured and operated. Anyway, as *similar* stimuli are presented, previous unsuccessful responses are substituted by successful ones. Learning is, then, a substitutive and continuous process.

Three major variables have been shown to account for the major variance in learning, namely the delay between stimulus, action and reward, the amount of reward, and the internal drive of the organism to value the reward (Commons and Liu, 2017). The characteristics of the task, the circumstances surrounding its presentation, and the circumstances of the individual who is solving the task also impact learning and performance (Duran et al., 2018). Learning modelers thrive on explaining and modeling how these variables interact and ultimately how they account for understanding and predicting learning abilities. However, while these factors explain the acquisition of a certain set of problem-solving skills, they are not sufficient to explain how organisms form many sets of problem solving skills, neither how organisms evolve from performing simple actions to more complex actions, as a result of maturation.

In the connectionist framework, the goal is to decrease the error as similar inputs are presented, such that the output generated by the model matches the desired output (Rojas, 1996). Learning is represented by an iterative weight update, in which weights change over time and substitute old, non-adaptive ones.

> During this process, as training inputs are fed to the system, the system goes from failing to succeeding. This might constitute a confusing aspect between learning and development in connectionist models, and other models in general.

Specifically, development implies that new patterns of perception and corresponding patterns of action emerge, without substituting for old ones. The Model of Hierarchical Complexity (MHC) introduces an important discrimination between horizontal and hierarchical complexity, which allows for discriminating learning and developmental processes, respectively. Horizontal complexity of a task refers to how many elements in chain compose the task, increasing its load. The horizontal complexity of the task might make it more effortful to solve, but the difficulty of its operations does not change. Differently, the hierarchical complexity of a task corresponds to it's a priori difficulty (Commons and Pekker, 2008).

Developmental modelers offer models that describe how these new patterns emerge and progress by stages, or plateaus, in a sequential manner (Dawson et al., 2003). Even in the presence of new patterns, or stages, the organism maintains the ability to move down to more elementary levels of information processing and performance, if the context so requires.

Hence, development is cumulative and discontinuous, imbedding the capacity to move up to complex levels, as well as the "capacity to move down to elementary levels, which provides enormous flexibility for intelligent adaptation" (Fischer, 2008).

Simulating cognitive development in connectionist models has been most closely achieved by generative architectures, such as the CC algorithm applied to problem-solving in the balance scale simulations (Dandurand and Shultz, 2009; Fahlman and Lebiere, 1990; Shultz et al., 1994). In general, generative architectures allow for the structure of the model to grow as the problem becomes more demanding, ascribing more complex information processing capabilities to the artificial system. Also, generative architectures allow for conciliating the continuous nature of learning and the discontinuous nature of development (Dandurand and Shultz, 2009; Shultz et al., 1995, 1994).

Yet, they do not allow for the system to move down to more elementary levels of information processing, as the new added elements (either units or connections) do not deactivate as simpler problems are presented.

This is the reason why the generative concept and procedure, in this perspective, should be revised.

3.2. Complexity in Developmental Models and Connectionist Models

So far, current simulations of the balance scale test are based on an assessment method that lies upon the notion that cognition is an information processing system. The goal of current simulations is to reproduce what is behaviorally observed — some balance scale problems are solved before others. The same goal is present in every attempt to simulate developmental problem-solving: simpler problems should be learned before more complex ones.

Yet, deducing that a reproduction of what is observed from the outside consequently reproduces what is ongoing inside is fallible.

Reproducing descriptions does not account for understanding the mechanisms serving behavior (Rijn et al., 2003). This is why modelers should avoid using mentalistic approaches, which make rules explicit and fundamental for problem-solving.

Before aiming at building a system that solves problems in a certain sequence, the goal should be to understand how a connectionist model represents developmental properties of problems, namely hierarchical complexity shown to apply to all domains of knowledge. If a transduction of problem-solving abilities is aimed at, then, modelers should concentrate on ways to conciliate the properties of developmental problems with the properties of artificial systems themselves (Elman, 1993), and allow for some room to the system itself finding the way to represent what it needs to represent. Connectionist models are the models of reference for such a task, as they are distributed representation models par excellence.

4. What Next

In order to represent the OHC, it is of the utmost importance to coordinate the continuous and substitutive nature of learning with the discontinuous and cumulative nature of development. This can be achieved by a generative architecture. A generative architecture allows for the emergence of higher-order connectionist patterns that solve more complex problems. An important feature of emergent patterns is that they will always correspond to U-shaped performance curves, where learning is taking place. At each new pattern, learning of the new components begins from scratch.

The drawback is that OHC has been attributed by the MHC to the number of input dimensions, type of operations, and output generation. This property of problems (their complexity) needs to be further clarified in interaction with the properties of connectionist models, namely inputs, structure, and type of desired outputs. In methodological terms, emergent connectionist structures should rely upon experiments that search for a compromise between the properties of the natural developing system and the properties of artificial learning systems. Another problem is how does the transition from one OHC to the next is processed.

> The next step is, then, to understand how an artificial learning model represents the OHC of the problems to solve and operations to conduct, and how does a transition between adjacent orders occurs.

Section C

Developmental Cognition in Modular Neural Networks: Stage Transitions are not explained by Hierarchical Integration

The simulation results exposed in this section has been orally presented at the 51st Society for Mathematical Psychology &16th International Conference on Cognitive Modelling, July 2018, by Sofia Leite and Pedro Pereira Rodrigues, entitled as "Simulating Developmental Cognition: Learning by Order of Complexity in Modular Neural Networks"

Developmental psychology has increasingly been taken into consideration in the design of algorithms that aim to learn in a developmental way. According to stage theories in developmental psychology, individuals improve and increase their problem-solving abilities as they progress through what have been called stages of development.

> A stage is characterized by performance invariance, whereas stage transition is characterized by unstable performance, due to the alternation between previous and emergent problem-solving capabilities (Dawson et al., 2003).

As mentioned, the Model of Hierarchical Complexity (MHC) (Commons and Pekker, 2008) is a developmental general stage theory of human behavior that postulates that stages of development are characterized by an Order of Hierarchical Complexity (OHC). The higher the OHC (or stage), the more hierarchically complex problems individuals can solve. However, in what concerns simulations of stages of development, it is yet to be determined how a connectionist model represents successive OHC, as the OHC has been unclearly attributed to the number of input dimensions, type of performed operations, and type and number of generated outputs.

In terms of stage transitions, although developmental theories (Case, 1987; Demetriou and Valanides, 1998; Fischer, 1980), including the MHC, do not clarify the mechanism in detail, it is postulated that "lower-order actions become the objects of higher-order actions"

where actions are a computational synonym of outputs. This mechanism is called Hierarchical Integration (HI), or Hierarchical Organization of Information. The present experiment is a preliminary simulation that aims to test the validity of the mechanism of HI for stage transition, using Modular Neural Networks (MNN). The following subsection briefly justifies why MNN are the most suitable connectionist architectures to use in the present case. Afterwards, the hypothesis on the validity of hierarchical integration is laid out.

1. Modular Neural Networks

Modular Neural Networks are composite structures of neural network models, in which each neural network is a module of the global system and works as an elementary unit. In these structures, the first stack is composed of several neural network classifiers operating independently and in parallel. The second stack weights the outputs of each of the previous models to produce a second-order output (Ting and Witten, 1999). MNN were chosen for the present simulation due to 1) the isomorphism between a MNN global structure and the structure of modules that represent hierarchical integration (Figure 4), and 2) the fact that HI postulates that lower-order outputs feed networks at higher-order stacks. MNN for hierarchical processing prevents the global model's complexity from growing with the complexity of the problem, allows data fusion from different sources, and allows for scalability (Rojas, 1996).

This option is commonly used to overcome the limitations of individual component networks when there is insufficient training data, when the training data carries a lot of noise or when it is highly expected that the learning algorithm will not find the optimal solution (Dasaratha, 1996; Ting and Witten, 1999). Specifically, MNN were initially used as ensemble techniques (Zhou et al., 2002), feature extraction techniques (Wang et al., 1998), and multi-class classifiers (Anand et al., 1995). In the first case, as ensemble techniques, each network in the first stack solves the entire problem. These networks might be set with different or similar parameters (Furtuna et al., 2012; Piuleac et al., 2010). Because the learning procedure is initiated with random weights and eventually with different parameters, each network will output different results. The network in the second stack weights the previous results [from each network] and attributes the greatest weight to the first-order network with best result. In the second case, for feature extraction, networks typically output different features, sometimes overlapping features, as they learn to represent the problem with different internal parameters. These first-stack features are then combined in the second-stack network, which receives a much more detailed representation of the problem and, consequently, outputs a better result than if no features had been extracted first. In the third case, multi-class classifiers, first-stack networks are used as lower-number-class classifiers, usually two-class classifiers. A second-order network weights the classifications of first-order classes. This third case is similar to the first case, but the difference is that, here, each first-stack network has its output range limited to a lower number of classes.

MNN have been applied to statistically neutral problems, where the chance of each class has a probability of 0.5 (Ghorbani and Owrangh, 2001), polymerization processes and polymer resin development (Fernandes and Lona, 2005; Zhang et al., 1997), time series forecasting (Leon and Zaharia, 2010) parsing (Irsoy and Cardie, 2014), natural language processing, sentence level sentiment analysis (Dong et al., 2014), and complex logical problems (Mengistu et al., 2016). This latter is an important experiment, which shows that a system evolves hierarchically to reduce the cost of connections, which simultaneously confers on it an enhanced ability to adapt.

Hence, MNN are not only models isomorphic to the proposed shape of hierarchical integration across modules, according to the MHC, but they appear to also approximate with biological plausibility the structural growth of a developing system (Mengistu et al., 2016).

For instance, each new *generated* stack in a MNN architecture would correspond to the emergence of a new higher-order stage, coordinating the information at lower-order modules. Nonetheless, MNN have not yet been applied to the problem of simulating cognitive development in a task such as the present one, but only to solving increasingly complex logical problems.

1.1. MNN applied to cognitive developmental problems: the balance scale problem

According to the MHC's current definition of HI and according to the regular use of MNN for hierarchical complexity processing, a stacked modular architecture succeeds if it is composed of two stacks. Stack 1 is composed of two unconnected modules or networks, each solving weight and distance problems, respectively. Stack 2 is composed of one neural network module that solves the range of conflict problems (Figure 9). Stack 2 receives lower-stack outputs as inputs.

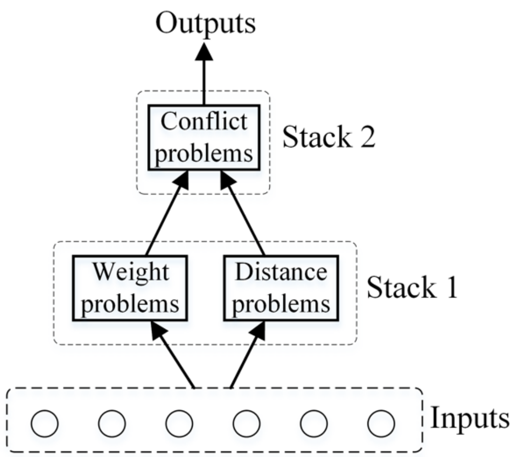

Figure 9 — Provisional structure of MNN applied to the Balance Scale Test

However, before an output is generated to solve a problem, the problem needs to be *perceived*. Hence, for outputs of a certain OHC to be generated, the problem needs to be perceived with that same OHC.If outputs are sequentially combined, as postulated by HI, in the limit, the lowest-order perceptual configuration of the problem leads to the generation of the highest order outputs, which is inconsistent.

> The main premise guiding this simulation is that "the complexity of an action is in accordance with the perceived complexity of a problem".

It is hypothesized that lower-order percepts become object of higher-order percepts, which will, in turn, trigger correspondent-order actions. Hence, stack 2, where weight and distance become coordinated, would receive another configuration of the original inputs of the problem. These are here called percepts. A percept is an environmental feature that needs to be perceived such that the problem is solved. In this case, the percepts would consist of the weight and distance on each side, each coordinated by the respective operation, which would be either sum (order 2 problems) or multiplication (order 3 problems).

2. Method

This section will describe how the data were created to simulate the learning context of the balance scale test, and the neural networks models that were created per module.

2.1. The Balance Scale Test simulation

Data representing all possible configurations of the balance scale test were simulated, with weight and distance values ranging from 1 to 15. Balance Scale configurations are represented as 4 integer-element input vectors [$weight_{right}$, $weight_{left}$, $distance_{right}$, $distance_{left}$]. Outputs ranged from [-1,1], where -1 represented the beam falling left, 0 represented the beam balancing, and 1 represented the beam falling right. The cut-off always remained 0.5 and -0.5 for discriminating the possible states.

The problem space was partitioned into simple and conflict problems. Simple problems are those where only weight or distance vary, solved by children at concrete stage 9. In conflict problems both weight and distance vary. Conflict problems were further partitioned into three classes – Conflict I, Conflict II, and Conflict TD. Conflict I problems are those where both dimensions indicate that the balance will fall to the same side. Hence, concrete stage 9 children are able to correctly predict the tipping side. Conflict II and Conflict TD problems are those where each dimension indicates a different side tipping down. In Conflict II problems, the sum of weight and distance at each side generates a successful response. Children at the abstract stage 10 can solve this subset of problems. In Conflict TD problems, only the multiplication of weight by distance on the same side solves the problem successfully, which requires that the child is at the formal stage 11. Hence, simple problems are order-1 problems, conflict II are order-2 problems, and conflict TD are order-3 problems.

2.2. Modular Neural Network (MNN) models

A control MNN composed of two stacks was created (Figure 9). The first stack was composed of two parallel unconnected neural network modules, one for solving weight problems and the other for solving distance problems. The second stack was composed of one neural network module for solving conflict problems. Experimental networks were also created, consisting of alternative neural networks for the second stack of the control MNN. Experimental networks differed in the input sets, so as to evaluate the importance of lower-order outputs in the formation of higher-order outputs.

Three different input sets were experimented. The second-stack network of the control MNN received lower-stack outputs, as postulated by hierarchical integration. Experimental network 1 received lower-stack outputs, as well as weight and distance of each side coordinated by sum. Experimental network 2 received only weight and distance of each side coordinated by sum. Experimental network 3 received weight and distance of each side coordinated by multiplication. This is illustrated in Table 4 and represented in Figure 10.

Table 4 — Inputs per order of problem complexity and respective examples

		Problem Type	**Input Stack 1**	**Input Stack 2**		
			[w_{left}, w_{right}, d_{left}, d_{right}]	Output Stack1	Left-side operation	Right-side operation
Control Network	**Stack 1**	Simple	[8, 8, 9, 6]	[0]	—	—
		Simple	[8, 8, 9, 6]	[-1]	—	—
	Stack 2	Conflict	—	[0, -1]	—	—
Experimental 1		Conflict	—	[0, -1]	[8+9]	[8+6]
Experimental 2		Conflict	—	—	[8+9]	[8+6]
Experimental 3		Conflict	—	—	[8x9]	[8x6]

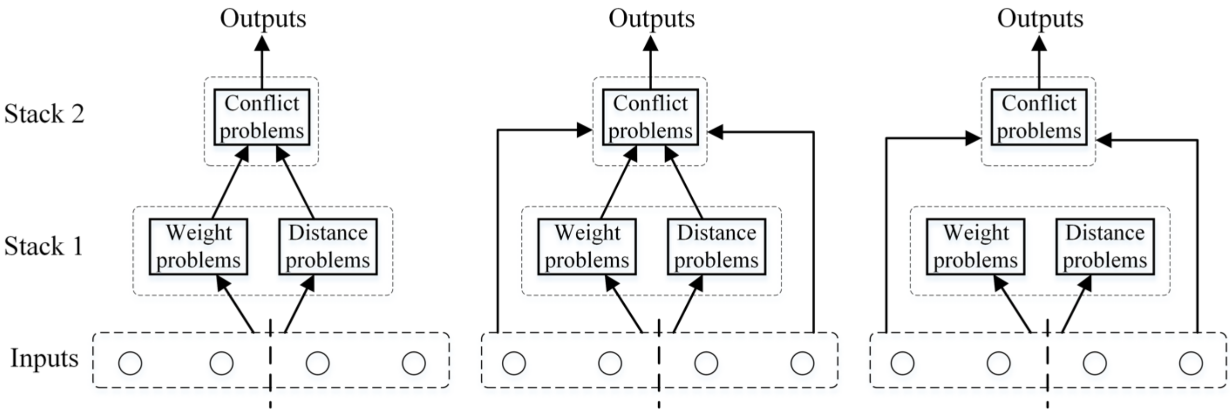

Figure 10 — Illustration of Control Network (left), Experimental Network 1 (center), and Experimental Networks 2 and 3 (right). Side-arrows represent left-side and right-side operations. In the right-side figure, they represent sum and multiplication operations.

2.2.1. Input data for neural network models

Stack 1 networks were trained alone and independently of each other. The training set consisted of weights and distances ranging from 1 to 5, which resulted in 5 balance problems, 10 weight problems and 10 distance problems. Each training routine was repeated 10 times for each network. The network repetition with better accuracy and lowest total error was kept for the test routine. The test set consisted of weights and distances ranging from 6 to 15. This resulted in 100 input configurations for the networks in the first stack and 100 generated outputs by each network.

At the second stack, these outputs were combined to produce second-order inputs, resulted in a total number of possible input configurations of 100^2. The correspondent set of weight and distance inputs were also combined, resulting in 100^2 original problem configurations. Afterwards, outputs were respectively concatenated to the input vectors, creating a 6-element input vector [$weight_{right}$, $weight_{left}$, $output_{weight}$, $distance_{right}$, $distance_{left}$, $output_{distance}$].

Of these combined configurations, balance, weight and distance problems summed a total of 1900 cases, which were excluded from the second-order dataset. The remaining 8100 configurations were conflict problems. Among these, 4050 were Conflict I, 3360 were Conflict II, which were joined in a single class of conflict problems (7410 cases) and 790 were Conflict TD. Both orders of complexity problems were then split into training (80%) and testing subsets (20%), controlling the representativeness of each class (falling left, balancing, or falling right) in all subsets (Table 5). Conflict-II problems were fed to experimental networks 1 and 2, whereas conflict TD problems were fed to experimental networks 3.

Table 5 — Number of input cases

	Order-2	**Order-3**
# of Training cases	5625	450
# of Test cases	1875	150

2.2.2. Network properties and functions

Neural networks were programmed in R language, using the "neuralnet" package (Fritsch et al., 2016). The following network characteristics remained constant throughout the entire experiment (Table 6). The learning rate and threshold were determined after experimentation.

Table 6 — Network properties and functions

# of Internal layer	Ranging from 1 to 2
# of Units	Ranging from 1 to 4 per internal layer
Learning rate	0.05
Threshold	0.15
Learning algorithms	Resilient backpropagation with weight backtracking (Igel and Husken, 2003)
Activation function	Logistic function
Error function	Sum of squared errors
Initialization weights	Randomized in the interval [-0,1; 0,1]
Repetitions	10

3. Results

Results are divided in two sections. Results concerning the Control MNN are presented first and results concerning the three Experimental networks are presented next.

3.1. Control network

In stack 1, both networks learned with total accuracy and classified 100% successfully the test set, requiring 1 unit in the internal layer. Also, both networks used approximately the same number of learning epochs: 80 for the distance network and 92 for the weight network (Figure 11 and Figure 12, respectively). The connection weights from the input to the internal unit are symmetrical in both networks, showing that each side is encoded symmetrically.

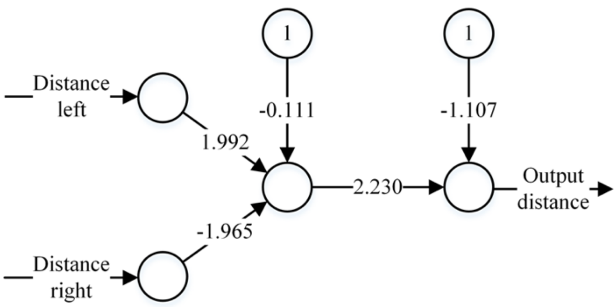

Figure 11 — Distance Network (Stack 1)

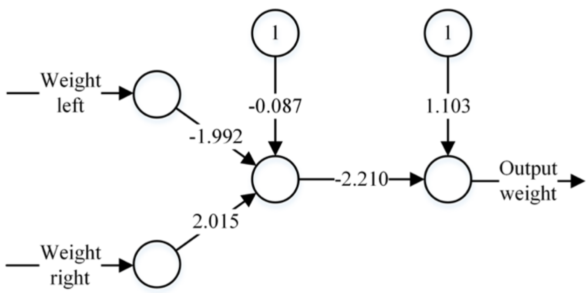

Figure 12 — Weight Network (Stack 1)

Stack 2 network only received the outputs of stack 1 networks. It did not learn to solve conflict problems. If order-2 and order-3 problems are classified altogether, the network performs with an accuracy of approximately 51%, independently of the number of units in the internal layer, and independently of whether it has one or two internal layers. The test set was classified with an accuracy of 52%, where approximately half of each problem order failed the correct classification.

This level of accuracy be considered to be at the chance level. This fact indicates that both orders were solved indiscriminately at the second stack, suggesting that either lower-stack outputs are non-informative for higher-order problem solving, or that the artificial model does not allow for the discrimination of hierarchical complexity of problems, or both. This will be clarified by the use of the experimental networks.

3.2. Experimental networks

Experimental networks 1 and 2 (used for solving order-2 problems), learned with an accuracy of 92,5% with only one unit in the internal layer.

Figure 13 depicts experimental network 1. If the connection weights are inspected, it is confirmed that lower-stack outputs are non-informative for information processing. The connection weights between lower-stack outputs and the internal unit are approximately the same (1.268 and 1.238), which means they do not convey discriminated information for problem-solving. Only experimental network 1 is depicted to explicit the null impact of lower-order outputs. Experimental network 2 only received right-side and left-side summed inputs, using all the information encoded in its connection weights.

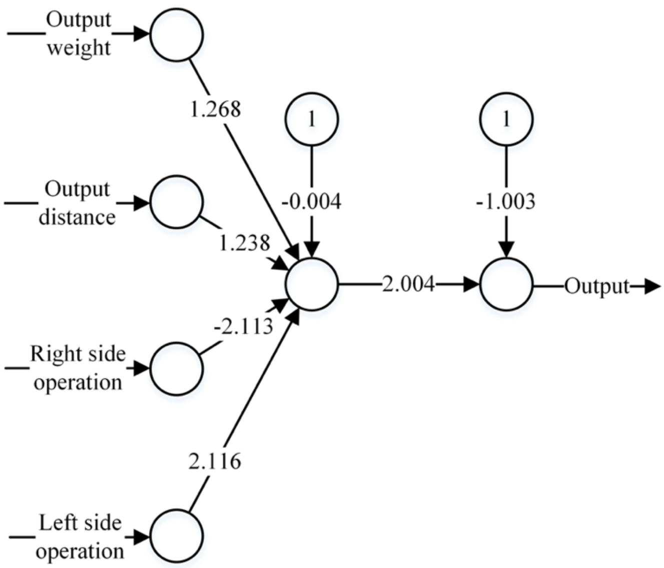

Figure 13 — Experimental Network 1 (Stack 2)

A fundamental difference between the two networks stands out. Whereas experimental network 1 required 32147 steps to learn and 16.77 seconds, experimental network 2 required 2262 steps and 1.01 seconds. This indicates that considering actions of the previous level accounts for no performance gain and for approximately 15 times more steps and time resources, given the parameters and network functions specified above (see section 2.2.2).

Also in both networks, the test set was classified with the same accuracy of 92,5%, where the failed 8% of problems were all conflict TD problems, of order-3 complexity. This is quite an obvious result, as order-3 problems require multiplication among the inputs and these networks received summed inputs. Yet, it is also interesting to note that, in the present case, the network did not use its distributed information processing capabilities to map a problem poorly represented at the input set, even with 2 layers. Rather, its performance showed an all-or-nothing approach. This confirms that, at least in this particular problem, the operations required to coordinate the inputs are fundamental for learning. The sum operation is specific of the second-order problem solving of stage 10 children.

Experimental network 3, which received only multiplied inputs, learned and performed with an accuracy of 100%, correctly classifying all cases. Experimental network 3 only used 1 unit in the internal layer, found symmetrical connections weights between inputs and the internal unit, required 3382 steps to learn and a minimum of 1.88 seconds.

4. Discussion

General-domain theories of behavioral development posit that behavioral complexity progresses along a sequence of stages and that transition across stages of development occurs through a process of hierarchical integration (HI) (Case, 1987; Commons and Pekker, 2008; Dawson et al., 2003; Demetriou and Valanides, 1998). HI means that stages are built out of one another, by a successive coordination of outputs. The present work served to test the currently accepted definition of HI in a connectionist computational model with a modular architecture. We further proposed another mechanism for stage transition, or stack transition. The proposed hypothesis was based on the idea that a cognitive task requires an initial perceptual appraisal of the problem for a subsequent recruitment of problem-solving resources and strategies.

It was proposed that HI is rather a mechanism in which lower-order inputs become object of higher-order inputs that generate correspondent-order actions than one in which lower-order outputs become object of higher-order outputs.

In sum, inputs, or percepts, should be reorganized into higher-order percepts, by means of certain operations, before a correspondent complexity output is generated.

Results confirmed the hypothesis. It is actually plausible to assume that the operations an individual uses to (re)organize the information are what matter for problem-solving. Furthermore, it is necessary to mention that a certain operation can only be performed when two circumstances have been achieved: 1) the problem has been perceived with a certain complexity; and 2) the individual matured enough to apply a certain operation to the problem. For instance, consider a child that looks at a balance with a Torque Difference configuration. If the child is still performing at the abstract stage 10, the problem will be solved based on the child's current resources — summing weight and distance on both sides. The output will be incorrect. Given this situation, the observer cannot attest that the child perceived the problem as a formal one, but lacks the ability to solve it accordingly. Rather, it is more congruent to assume that the child, whether in the face of a simple or complex configuration, is only able to perceive it as an abstract one, according to the perceptive and cognitive skills acquired so far.

These results have important implications for the formalization of the mechanism underlying stage transition, as well as for the simulation of stage transitions in an artificial learning system. The fact that an output that correctly solves a certain complexity task is necessarily of the same order of complexity as the task it solves, called "task-action", is maintained (Commons and Pekker, 2008). Yet, the way the model postulates how higher-order outputs are formed has been challenged.

> It is the internal operations that matter, in combination with problem dimensions.

Once the mechanism of HI has been invalidated, the use of Modular Neural Networks has similarly been disregarded for the purpose of reproducing stages of development in an artificial model. In the present example, the increasing complexity of operations required for solving the balance scale test imply that children go from the ability to count (in simple order-1 problems) to the ability to sum (in conflict problems of order-2), and from the ability to sum to the ability to multiply (in conflict problems of order-3). Interestingly, these three arithmetic operations are assumed to be recursive.

5. What Next

An important question that drives future work is: how does a neural network model represent increasing OHC?

An answer to this question will allow for the identification of changes in a connectionist structure, as new developmental capabilities emerge. In turn, these changes will hopefully inform how more complex capabilities can be generated in a developmental algorithm. Ideally, this will not only apply to the development of a neural network model that solves the balance scale test as children do, but also to other developmental domains.

Section D

Neural Correlates of PostFormal Stages of Reasoning: Biological Determinants of Developmental Stage

The original version of the section that follows has been published as "Leite, S., Barker, C.D., Lucas, M.G., 2016. Neural Correlates of Postformal Stages of Reasoning: Biological Determinants of Developmental Stage. Behavioral Developmental Bulletin. Vol. 21, pp. 33–43 DOI:10.1037/bdb0000012. The following text includes some modifications, as well as some excerpts that have been added

Post-Piagetian theories of development claim for the hierarchical development of reasoning abilities throughout life. A sequence of ordered stages is usually defined, even though the processes that underlie stage transition are debated.

Particularly for domain-general theories of development, stages of development are accepted to hold some sort of performance invariance. Performance invariance, or stability, is what imposes constraints on flexibility.

Within the scope of the present dissertation, performance invariance has been previously (Chapter II, Section A) discriminated as hierarchical and horizontal invariance, imposing different types of constraints on flexibility. Hierarchical performance invariance, or hierarchical stability, is attributed to the stage of development, and allows for flexibility within the boundaries of task complexity. Hierarchical performance invariance, i.e., stages and stage transitions, is the focus of the present work.

The Model of Hierarchical Complexity (MHC) is one of these Post-Piagetian theories, which formulates that stages are defined by an Order of Hierarchical Complexity (OHC) and demonstrates that the OHC explains 99% of observed behavior (Giri et al., 2014). Thus, the MHC has a high predictive power when applied to experimental settings. Theoretically, it has also provided for a great degree of understanding of human behavior as its formulations reveal that cognition is primarily a product of a structural property of organisms. However, the MHC, as other behavioral theories, are unclear in the definition of the underlying mechanisms of stage transition. The lack of clarity in regards to the representation of stages and stage transitions at a neurocognitive level leads to questioning what are the biological parameters that account for hierarchical performance invariance (hierarchical stability) and bound hierarchical flexibility. This is important for understanding the underpinnings of development in the neural architecture, as well as for informing how stage transitions can be implemented in a developing connectionist algorithm.

> To ask whether the biological architecture works as a biological boundary for performance flexibility and stability is synonymous to ask whether the OHC can be represented in the brain.

In this sense, the focus is now on the contribution of developmental cognitive neuroscience for the understanding of growth in hierarchical complexity of cognitive capacity in the neural architecture. Recently published results suggest that the neural architecture actually supports the hierarchical growth in stage, both developmentally and evolutionarily. Combined with insights from developmental cognition and developmental cognitive neuroscience, this section aims at establishing a biological plausible ground for understanding how stage might be traced in the brain and which are the structures and functions that determine the growth in stage at a neuro-cognitive level.

Furthermore, the MHC has yet uncovered why some individuals seem to be hardwired differently, leading to differences in stage of performance and, consequently, in behavioral patterns.

> The work here presented is a proposal for 1) identifying what changes in the neural signature as higher stages are achieved and 2) ultimately answering why, not how, some individuals achieve higher developmental stages than others.

The proposal is to look for this answer through the neural correlates underlying stage of performance, namely through power spectra electro cortical activity (sEEG) and neuroimaging correlates (MRI). This proposal lies in the overlap between developmental psychology and developmental cognitive neuroscience and aims to provide for acknowledgement of the biological basis of higher-order cognition, both at a structural and at a developmental level. Answers to these questions will also further improve the predictive power of the MHC, and represent an important input for simulating human cognitive development in an artificial intelligence algorithm.

This work is divided into four subsections. First, the axioms of the MHC will be briefly described and how they apply to predicting behavior; second, some literature findings on the neural correlates of intelligence and cognitive development will be revised; third, some methodological considerations, specifically in what refers to data analysis, will be presented; Finally, a discussion will be elaborated on how the expected outputs can contribute to improving the prediction capabilities of the MHC and how they fit in the development of an algorithm that pretends to simulate human cognitive development. Given the scope of application of the expected outputs, although this specific study concerns the field of developmental cognitive neuroscience and is defined within the bounds of a neuroscientific study, it is included in a wider line of research concerning computational modeling, developmental psychology, and behavioral prediction, all together harvesting for a stronger theoretical construction.

1. Biological Underpinnings of Stage of Development

"Smartness" is defined as the ability to solve problems or tasks, which are measured by an Order of Hierarchical Complexity (OHC) (Commons and Pekker, 2008). Complexity is operationalized as the number of concatenation operations a task contains. At each order of complexity (or at each stage), the individual is able to perform the correspondent complexity actions and solve the correspondent complexity tasks. Higher-order task-actions are characterized by the non-arbitrary coordination of lower-order task-actions. The individual becomes capable of organizing and combining immediately lower order actions in a non-arbitrary way. Non-arbitrariness is the property that imbeds new configurations with meaning. The discovery of the OHC as the strongest predictor of behavior across domains (Commons et al., 2014b) turned the MHC into a non-mentalist structure-driven approach. This means that the MHC holds its validity independently of the mental strategies for problem-solving. This also means that the existence of a supporting biological structure for cognitive development is under proof.

Different from Piagetian perspectives, the MHC presents a conception of intelligence and development that goes beyond formal operations, in that cognitive development is a functional mechanism that pervades throughout life (Commons et al., 2014a; Commons and Pekker, 2008). This post-Piagetian conception of human development throughout adulthood simultaneously constitutes the major strength and weakness of the model. The major strength because the amount of evidence collected so far models the shared properties of development, showing that there is a universal developmental sequence for all organisms (Commons et al., 2014a). The major weakness comes from the fact that the model has not yet explained the reason underlying inter-individual variability. In other words, why some individuals attain higher stages than others. The present method also aims at ultimately answering this remaining question: what limits stage?

> The MHC provided recent evidence favoring a biological perspective over stage of development.

The first is that stage holds across domains (Commons et al., 2014a; Giri et al., 2014), which suggests a general activation mapping or structure in the brain that supports cognitive performance in all domains, even though concept formation and representation has been consistently shown to activate domain-specific regions (Bauer and Just, 2015). Second, it shows that stage develops as a function of $\log_2$(age), which suggests that the roots of stage achievements are ontogenic (Commons et al., 2014c). This finding has been supported by cognitive development literature (Wendelken et al., 2015). Third, the MHC showed that a power function models the increase in the number of neurons as the highest stage of a species increases, with $r = .874$ (Harringan and Commons, 2014), while further work favors of a view of cognitive abilities that is centered on absolute numbers of neurons (Herculano-Houzel, 2009). This evidence does not directly provide an argument for stage specific differences in humans, but hints that cognitive capabilities are traced back in the neural architecture across species. Fourth, there is intra-species evidence of behavioral development going together with a dynamic growth of neuronal connections (Qin et al., 2014). Hence, two valuable premises can be traced. One the one hand, a fixed number of neurons is correlated with the mean stage that a species achieves, being it the stage of formal operations in humans (Commons et al., 2014b). On the other hand, the number of connections changes throughout development, which points towards a dynamic adjustment of the neural architecture within some fixed anatomical parameters, as experience proceeds (Qin et al., 2014). Taken these evidences together, the MHC suggests that cognitive development is basically dependent upon these structural and functional biological correlates, providing an argument for biology controlling stage.

1.1. Inter-stage differences

Several brain-based indicators provide evidence that there is a common ground between brain dynamics and spurts of cognitive development, which follow positive correlations as children grow up. Namely, the number of neurons and synapses, brain mass, myelination, patterns of brain electrical activity, cortical thickness, skull size, all represent a partial brain-based description of cognitive development (Fischer, 2008; Hudspeth and Pribram, 1990). Discontinuities are evident in many of these brain indicators. For instance, developmental transitions seem to be accompanied by an overabundance of synapses. Afterwards, depending upon experience, some synapses are pruned-back and others are strengthened. These processes have been modelled based on two fundamental curves — one is an inverted U-shaped curve that represents the rapid increase and decrease of synapses, where the pruning phase corresponds to skill acquisition. These inverted U-shaped changes occur at different timespans, depending on the cortical region (Casey et al., 2005; Morita et al., 2016). For instance, in the visual cortex, synaptic activity rapidly increases around 2–3 months, achieves a maximum at 4–12 months, and then decreases to the level found in adults around 2–4 years. Differently, in the prefrontal cortex, the synaptic activity similar to that of adults is achieved at the age of approximately 15–20 years (Morita et al., 2016).

There is also a linear linear increase with age of the volume of the white matter, which continues until the age of approximately 20 years old, as well, in all brain regions. This linear increase in white matter corresponds to myelination of those axons that remain active, after synapses are pruned back (Morita et al., 2016). Further experiments were conducted showing that gray matter volume and intrinsic connectivity not only can explain, but can further predict performance gains, opposite to explicit behavioral measures, such as neuropsychological assessment scores. It was still demonstrated that neural correlates capture structural and functional changes as learning and skill acquisition occur, even if in a restricted timeframe of 8 weeks, and even if no stage transition occurs (Supekar et al., 2013). In general, experiments conducted to date found promising results to the identification of the neural signatures underlying learning, skill acquisition, and development, both with Magnetic Resonance Imaging (MRI) (Cho et al., 2012; McClelland et al., 1995; Qin et al., 2014) and with spectral content of electroencephalography (sEEG) data (Fischer, 2008; Hudspeth and Pribram, 1990; Klimesch, 1999).

In regards to MRI data, the majority of studies have been conducted regarding learning and skill acquisition. It has been suggested that skill acquisition in children is a phenomenon accompanied by a shifting from procedural-based strategies to retrieval-based strategies, and that this shifting is mainly associated with the hippocampus-neocortex system (Cho et al., 2012; McClelland et al., 1995). The authors further suggest that this shifting is held across different domains and, thus, that this neural system might be critical for cognitive development in general. A subsequent study was conducted showing that this shift from procedure-based strategies to memory-based ones goes along with a decreased activation in prefrontal regions and increased hippocampal activation. Beyond childhood, retrieval-strategy-use continued to improve through adolescence into adulthood and was associated with decreased activation, but more stable inter-problem representations in the hippocampus (Qin et al., 2014). A complete review of the role of the hippocampal-prefrontal system in learning and memory is supportive of these results (McClelland et al., 1995).

In regards to sEEG signatures, there is a considerable corpus of knowledge linking brain dynamics and cognitive maturation throughout life (Fischer, 2008). Both tonic and phasic measures of sEEG patterns have been linked to cognitive performance, with the latter being related to performance in problem solving. Relative energy shows systematic growth curves in the occipital-parietal regions, with this growth proceeding through spurts or plateaus, as is observed for cognitive development (Fischer, 2008). Based on these findings, the nested-network hypothesis was proposed, which considers that the emergence of cognitive levels correspond to a large cycle of growth of energy, coherence, and other brain measures. Curiously, increases in alpha energy occur through spurts until adolescence, where individuals are likely to be achieving the stage of formal operations, the mean stage for humans. After adolescence, and mainly between the ages of 60-80, the pattern reverts, showing a decrease in alpha energy. This has been suggested to be the results of interference of neurological degenerative conditions (Klimesch, 1999). Furthermore, changes in alpha and theta power also show a positive age-related correlation. Delta and Theta bands power decrease with age, while alpha increases. Interestingly, these changes are also consistent when comparing children without learning disabilities with children with learning disabilities or neurological disorders, pointing towards the relationship between the power bands and cognitive performance (Klimesch, 1999). Taken together these findings, the alpha band has been associated with cognitive performance, mainly speed of processing, memory (Fischer, 2008) and attention (Klimesch, 1999), as well as with general cognitive performance throughout life ((Fischer, 2008; Hudspeth and Pribram, 1990; Klimesch, 1999). During problem-solving, synchronization and desynchronization of alpha power have been studied. Lower-alpha desynchronization has been systematically assigned to reflect attentional resources during problem solving, upper-alpha has been linked to the processing of sensory-semantic information, whereas theta synchronization also appears to be correlated with working memory or episodic memory performance (Klimesch, 1999).

1.2. Inter-individual differences at each stage

Both MRI and sEEG data provide strong evidence towards a shared developmental path inscribed in the brain. However, the majority of existing studies overlook inter-individual differences in cognitive performance. Actually, a closer look to these studies reveals that inter-individual differences play an important role. For instance, although there is a common hippocampal-prefrontal connectivity pattern underlying learning and skill acquisition in several domains, irrespectively of individual differences, performance gains range from 8% to 198% (Supekar et al., 2013). Moreover, inter-individual variability plays as large an effect as that of age-related changes. Also, within the frequency bands, there is a high variability, too, in how to define sub-bands (Klimesch, 1999). This shows how variable performance gains might be under the same experimental circumstances, which, again, should not be overlooked.

In one relevant study addressing this issue of inter-individual variability (Lee et al., 2006), subjects' IQ was measured as a general cognitive capacity (g-capacity), splitting the sample into two groups — g-superior and g-average subjects. Tasks similar in shape but differing in g-loading were administered to each group of subjects, while their patterns of brain activation were measured through fMRI. A brain signature was found in both groups concerning bilateral activations in lateral prefrontal, anterior cingulate, and posterior parietal cortices. These g-task-related neural substrates were most likely to rely on the fronto-parietal network that was previously reported to constitute the neural bases for fluid reasoning and working memory (Lee et al., 2006). A brain signature was also found between groups, with the superior g-group showing much greater percent signal changes of the regions of interest than the average g-group. The most significant gap between groups was in the posterior parietal cortex. These findings are further supported in the literature by a recent result obtained from the Neurodevelopment of Reasoning Ability study (NORA) (Wendelken et al., 2015). It was confirmed the involvement of the Fronto-Parietal network in detecting differential reasoning abilities. The authors found an increased connectivity between the Rostro Lateral Pre Frontal Cortex and the Inferior Parietal Lobule in the mature reasoning system, in opposition to an immature neural system. The fronto-parietal network is also at play when differences are being measured for EEG power spectra. Namely, delta and theta frequency bands decrease in power with age, while alpha frequency band power increases, with this increase starting at posterior derivations and ending at more anterior recording sites (Klimesch, 1999), which is consistent with recent big data analysis (Taylor et al., 2015). Recently, fronto-parietal networks have been associated to higher-order cognitive functions majorly because they underlie the representation and management of concepts with the highest levels of abstraction (Taylor et al., 2015).

From the findings reported above, it is deduced that the hippocampal-prefrontal network is involved when considering longitudinal designs for functional MRI data, irrespectively of inter-individual differences. This network is involved in a gradual change in problem-solving strategies, from procedural to retrieval-based, which occur independently of the rate of learning. However, if the focus shifts to inter-individual variability of reasoning abilities and differential learning rates, the regions associated with differential activation are no longer observed in hippocampal activation. These are reported to rely on the fronto-parietal network instead (Lee et al., 2006; Wendelken et al., 2015), which is consistent for both fMRI and sEEG data.

2. Finding Stages in the Brain

Limitations of existing studies concern, first, the fact that these comprise unsystematic sets of tasks, which results in outcomes contaminated by task specific variance.

> In other words, they do not present a sequence of tasks with a priori measured difficulty or processing load. This does not allow to extrapolate the results with confidence to other domains.

Second, they restrict their object of analysis to an early period in life, mainly in neuroimaging studies. This impedes from taking conclusions in regards to the development of higher-order cognition and falls apart of the questions deemed to answer: how stages are represented in the brain and what limits stage? Third, all these brain-developmental processes compete and occur in parallel, accounting for the great plasticity and flexibility of brain functioning, which is mainly implemented in a network (Morita et al., 2016). The Interactive Specialization (IS) approach assumes that the specialization of a cortical region is interdependent with its neighbor regions and connection patterns. In other words, "the response properties of a cortical region are determined by its patterns of connectivity to other regions as well as by their own current activity" (Johnson, 2011).

> The emergence of a higher-order network to support a higher-order stage is consistent with viewing cognitive regions functioning in a jointly integrative manner (Kanwisher, 2010; Smith, 2005). As regions are integrated, the network grows in structural complexity.

In line with a complex system's perspective, the IS approach suggests that development and skill acquisition are counterpartyed by a reorganization of interactions between different structures and regions. "This re-organization process could even change how previously acquired cognitive functions are represented in the brain" (Johnson, 2011), which makes it more difficult to operationalize, observe, and simulate.

The current proposal is that stage of development is best understood as a global structure of dynamic processes and functions, rather than a set of identifiable elements. The method we devised aims at identifying 1) the global structure corresponding to each OHC, as well as 2) changes that occur in such structure as new stages emerge. A structure representing an OHC is responsible for attracting, or perceiving, a specific set of information from the environment, contains rules for problem-solving, and outputs a set of correspondent complexity actions. The emergence of a higher-order structure leads to perceiving the environment as a higher-order set of information, to processing information by higher-order rules, and to outputting a set of higher-order complexity actions, and so on and so forth until the maximum stage of development is reached and the maximum difficulty problems are solved. Furthermore, the variability in the highest stage to be attained will certainly be correlated with variable changes in structure as development occurs.

It is, then, proposed that a different canvas is necessary to uncover 1) the parameters that correlate with stage of development and 2) the factor that loads on stage variance, ultimately answering 3) what limits stage. The proposal is to look at inter-stage differences, as well as at inter-individual differences within stage. In fact, although development and brain maturation yield significant similarities across subjects, there is also strong evidence of individuals displaying different rates of development (Commons et al., 2014b) and different mosaics of a developmental path (Abellán et al., 2015), not to mention the differences that show up in populations with disabilities and/or neurological disorders. Hence, the proposed methodology hopefully allows for a comparison between neural activations of subjects who perform at different stages of development. This will allow for extracting not only the shared properties of problem solving in the brain, with problems, solutions, and competence operationalized by an OHC, but also to extract the differentiators. A set of regions of interest and EEG features are addressed to suggest where to look and which data to analyze. Namely, the percent signal activation in regions of interest (fMRI) and power and energy of frequency bands (sEEG).

In order to model inter-individual variability from these data, it is further proposed to extract the Saturation Index (SI) for each physiological measure, which is claimed to represent the processing load along the developmental path of subjects. Ultimately, it is expected that the SI of each measure will be correlated and a general SI can be extracted from it. SI is then a within-subjects measure that intends to model the individual dynamics of development. It is expected that the potential to achieve a certain developmental stage can be characterized by a specific SI, as if the SI is a dynamic neural signature underlying, or carrying, development. Basically, the hypothesis is that the SI will face a faster relative increase for potentially lower stage subjects than for potentially higher stage subjects, reflecting that an increase in task complexity requires higher cognitive resources for lower stage subjects than for higher stage subjects. The SI is, in fact, closely related to the functional meaning of the Index of Harmony

(IH) calculated for assessing and predicting developmental problems from birth to adolescence in a ten-domain general model of child development (Abellán et al., 2015). The idea of a IH, an individual index of development, is "fundamental to give independence to the comparison of individual development in relation to statistical norms, since it permits each case to be contrasted with itself." (Abellán et al., 2015). Further on, the SI is the numerical indicator that will allow this study to see continuation in different fields, such as artificial intelligence and behavioral prediction. In sum, the extraction of a SI is assumed as one of the major expected achievements of the present proposal, along with uncovering some regions that are differentially activated in the face of different stages of development, as is shown by some important literature findings.

2.1. Hypotheses

Eight hypotheses build on these findings. The first four hypotheses stand for characterizing patterns of brain activation across stages and tasks. For MRI data, it is expected that different stages will show up in the brain as differential patterns of activation in the Fronto-Parietal Network, namely in the Pre-Frontal Cortex (PFC) and in the Posterior Parietal Cortex (PPC). A positive correlation between activation in these regions and an increase in the complexity of tasks is expected (H1). It is also hypothesized that connectivity between PFC and PPC is positively correlated with stage (H2). Still, higher stage subjects show a decreased activation in the regions of interest when compared to lower stage subjects, when performing the same task, which order of complexity should be equal or lower than the order of lower stage subjects (H3). For sEEG data, tonic alpha power increases and theta decreases with the complexity of the task and that phasic changes also show a higher theta synchronization (H4). The remaining four hypotheses are concerned with modeling the SI. Fifth, for MRI data, higher stage individuals show a slower increase in activation in the regions of interest while solving tasks with increasing order of complexity, than do lower stage subjects (H5). Higher stage subjects show a more pronounced increase in the connectivity between the regions of interest that lower stage subjects (H6). For sEEG data, the increase in alpha and the decrease in theta power along increasingly complex tasks is significantly more pronounced in lower stage subjects than in higher stage subjects (H7). Finally, during problem-solving, theta synchronization is more pronounced along increasing complexity tasks in lower than in higher stage subject (H8), reflecting a more effortful working memory in the first group.

2.2. Goals and Objectives

In sequential order, it is the first objective to confirm existing findings by operationalizing reasoning abilities and stage of development as the order of hierarchical complexity. The first four hypotheses stand for characterizing mappings of brain activation, both with MRI and sEEG data, which are expected to confirm what has been shown in previous studies. The following objective is to answer why, not how, some individuals achieve higher developmental stages than others and to look for this answer through the neural correlates underlying stage of performance. This objective will be accomplished through calculating a Saturation Index that informs the progression of processing load along the problem solving of increasingly complex tasks. This will be done in the remaining four hypotheses of this study. It is still an objective to use the results of this study to improve the behavioral predictive MHC and to contribute for the development of a connectionist model that attempts to simulate the growth in complexity of reasoning abilities.

3. Method

The methodological section includes the description of dependent and independent variables to be considered, the proposed experimental design, and the data analysis rationale to conduct hypothesis testing.

3.1. Independent and Dependent Variables

Reasoning abilities will be operationalizable by assessing stage of development of participants, as is determined by the MHC. The MHC has a high predictive power when applied to behavioral analysis (Giri et al., 2014), it measures the difficulty of tasks to avoid the interference of task variance noise in data analysis, it is in agreement with further mathematical behavioral developmental models that correlate age and stage (Wendelken et al., 2015), and it has recently shown that the majority of intelligence tests fail to detect postformal capacity (Featherson et al., 2016). High IQ score probably represents Formal Stage and Systematic Stage 12 performance. The Low IQ represents Concrete Stage 9 and Abstract Stage 10 performance. IQ does not measure Metasystematic Stage 13; however, given the existing data on the progression of stage (Commons et al., 2014c) if an adolescent is performing at the Systematic Stage 12, it is highly likely that the person will move up to Metasystematic Stage 13 (Featherson et al., 2016).

If previous studies found consistent findings given the maturity of reasoning abilities, irrespectively of the operationalization criteria, this proposed study will also certainly find these differences, concluding that this methodology is sensitive enough to identify differences in stage. Stage of development will be assessed through instruments developed so far by the MHC. As dependent variables, we will measure MRI and sEEG correlates of task performance.

3.2. Experimental design

This study comprises a cross-sectional experimental design, where the scope of observations will be restricted to abstract, formal and postformal stages of development, namely systematic and meta-systematic. For modern humans, the range of stages in intact adults is from abstract stage 9 to postformal stages 11 and 12. The mean stage of performance has been shown to be the formal operational stage 10 (Commons et al., 2014b). Stages beyond formal operations (Stage 10), including systematic (Stage 11), metasystematic (Stage 12), paradigmatic (Stage 13), and two other very rare stages, have also been described by the Model of Hierarchical Complexity (Commons et al., 2014b). Hence, according to the MHC, inter-individual variability in attaining higher-order stages is only pronounced when we move up to formal stages and beyond. Subjects will be selected for the study based on their stage of development, irrespectively of educational background. Experimental groups will match in gender and age.

3.3. Hypotheses testing

Data analysis procedure will be based on Representation Similarity Analysis (Kriegeskorte et al., 2008). H1, H2, H3 and H4 are tested to confirm previous findings in what concerns the relationship between the increasing of reasoning abilities and the emergence of differential patterns of brain activation. From this confirmation, it will be demonstrated that operationalizing reasoning abilities as stage of development does not introduce an uncontrolled bias in posterior data analysis. In the following hypotheses H5, H6, H7 and H8, the objective it to model how mappings of brain activity (representation mappings) progress along the performance in increasingly complex tasks for a specific group of subjects. Afterwards, we representation mappings that show up during problem solving in consecutive complexity tasks will be compared; pair-wise comparison will be called transition mappings (Table 7). Representation mappings correspond to each cell on table 1 and transition mappings correspond to the arrows transiting from one cell to the other. Transition mappings underlie the functional meaning of the Saturation Index.

Table 7 — Representation mappings and transition mappings

		Experimental groups			
		Abstract	Formal	Systematic	Meta-Systematic
OHC of Tasks	Concrete				
	Abstract				
	Formal				
	Systematic				
	Meta-systematic				

4. Limitations

Two limitations are attached to this experimental design. First, one cannot know if the subjects who compose each experimental group have already attained their highest stage; one can only be aware that their neural architecture, when compared to matching age subjects, is higher. In order to overcome this limitation, one possibility is to set a lower age limit based on the evidence that stage progresses as $\log_2$(age) (Commons et al., 2014c);

however, this solution is not free from methodological problems, as setting a lower age limit will possibly introduce the interference of cognitive degeneration in more aged subjects, which might begin occurring at the age of 40 (Klimesch, 1999). The second limitation of this research proposal concerns the fact that one is not evaluating changes in the neural architecture *as* a stage transition occurs, which would be the ideal scenario, but only possible through capturing a once in a lifetime event. As such, observations are restricted to how the neural architecture changes in the face of different complexity tasks and assume that these changes, or adjustments, somehow remain after a new stage has been achieved, similarly to the remaining of a phylogenetic process of evolution and development pervading in the organism.

5. Application & Future Work

The question of what limits stage of development differentiates this study from others that have been conducted in the field of developmental psychology and cognitive neuroscience. This is a relevant topic of research that has never been addressed. In regards to the proposed method, besides serving the fields of psychology and cognitive neuroscience, it further serves other branches of application fields. For instance, the field of behavioral assessment and prediction and artificial intelligence.

5.1. Behavioral Prediction

The Model of Hierarchical Complexity is a behavioral assessment theory of development with a high predictive power ($r = .991$) (Giri et al., 2014). Mean stage is determined as the logarithmic function of age, hence, younger individuals attaining higher stages than their counterparts are assumed to achieve higher stages in the future. However, there is no clear predictive evidence of this fact nor there is evidence of a biological mechanism controlling stage. Because this study is proposed to result in the extraction of the SI – an index that informs about the highest stage to be achieved – it is closer to further improve the predictive capability of the MHC in what concerns later stages of life.

One specific area of application concerns the development of educational and pedagogic practices more adequate to people's general cognitive capacity. If the expected results are ultimately proven, the SI would be a quantitative indicator of differential educational strategies. With a new educational approach set up, the reverse research line can be initiated — whether adequate educational practices actually change the development of the brain along the suggested lines. The idea is similar to previous research on modeling general properties of development along with individual differences, with the goal of determining how the individual developmental path can be improved and optimized (Abellán et al., 2015). The same principle applies for hierarchical complexity measurements for assessing how and where employees best fit into organizations based on their task performance (Commons and Robinett, 2013). This is important because in many societies in the world, especially among certain sectors, there is a belief that there are no biological differences underlying how smart someone is. People associate differences with education and motivation. Hence, people who are not hardwired to achieve the highest stages are possibly treated unfairly because the expectations for them are unrealistic. If biological differences are found, it may inform interpretations of behaviors that support a more ethical and fair society.

A second area of application might also concern psychiatry and law. Results could add a new lens for verifying that observed arrested development of the interpersonal domain has hierarchically complex neural correlates of brain behavior that correspond to observed hierarchical complexity performance. This can inform both psychiatry and law in their respective efforts to adequately approach behavioral deficits and crime to two ends. The first is to improve and correct maladaptive or criminal behavior (Commons and Miller, 2011) by applying the above mentioned renewed educational practices. Once again, appropriate approaches could be informed according to the saturation index calculated for each individual. The second is to predict and prevent future criminal behavior. This would be based on the brain signatures of individuals who carry some neurological limitation and which might be cause, under certain conditions, to suggest a higher probability for social threat. For instance, people with Asperger syndrome are considered the most dangerous people because they show no social perspective-taking. The model for predicting crime would be multiplicative, with stage, neurological disorder and social perspective-taking interacting together and all being extracted from the brain. Representation mappings would inform about the relative threat people represent at present as they convey information about current stage of development; the SI would inform about the social threat they might represent in the future, given the stage they are hardwired to achieve.

5.2. Computational Cognition: The Hierarchical Stacked Neural Networks model

The MHC is also the grounding theory for a computational model of cognitive development, called "Hierarchical Stacked Neural Networks". This is a neural-networks algorithm that simulates successive behavioral hierarchical stages of development of individuals. There is an a priori conceptual isomorphism between the MHC and the algorithm, which stands for ascribing human developmental abilities to AI systems, not yet seen in AI field (Commons, 2008). In the artificial model, information flows continually from the lower order stack to the higher order stack, in the direction of increasing complexity. In order to do that, one needs to characterize stack transitions.

To simulate successive behavioral stages of development implies building a system's architecture in stacks. Each stack is the computational synonym of stage and consists of a neural network with a particular structure that makes it generating the hierarchical complex actions of the particular stage it simulates. In a neural networks model, the topology (number of units and connectivity pattern among them) and learning algorithm partially determine the learning and generalization capabilities, similarly to what happens in the brain.

> If stages of development are imprinted in the brain and the emergence of new stages correspond to the emergence of new patterns, then, each stack is the imprint of a stage in a connectionist model and each new stack will comprise the emergence of a new stage, or network structure.

In order to establish a parallelism with the present methodology, exposed for neuroscientific study, each stack accounts for representation mappings and stack transition accounts for transition mappings. The proposed method has, actually, the advantage of allowing that the same procedure is applied to both natural and artificial systems. This allows that conclusions across fields are constrained by each other and have likely more descriptive and explanatory power.

6. What Next

The primary goal of this proposed method is to find the biological correlates that describe, explain and limit the highest stage attained, which has never been addressed before. Existing studies mostly point towards the biological underpinnings of learning and skill acquisition. Fewer address the neural correlates of development. However, they carry some limitations, such as not controlling for stage of development, and lag behind addressing the crucial aspect of individual differences in performance gains and cognitive capacity. Those that address this issue are far from suggesting that biology might actually explain *and* limit cognitive capacity. It was proposed that representation and transition brain activation mappings were extracted from MRI and sEEG data, and that a Saturation Index (SI) was further calculated. This numerical indicator is expected to represent a predictor of the highest stage achieved, as the SI is intended to inform about the processing load that each increasingly complex task requires.

> The next step concerns the application of the proposed method to computational modeling — specifically to the development of a connectionist algorithm called "Hierarchical Stacked Neural Networks" model that aims to learn in a developmental way.

CHAPTER III

METHODOLOGICAL CONSIDERATIONS FOR SIMULATING COGNITIVE DEVELOPMENT IN AN ARTIFICIAL MODEL

Given the concepts and resulting ideas from Chapter II, the present Chapter III, exposes methodological considerations and procedures to identify how connectionist stacks represent orders of hierarchical complexity. It responds to the second objective of the present dissertation "Definition of a method for identifying structural changes in a developing connectionist system". All the methodological aspects and procedures here delineated will serve as the basis for the subsequent Chapter IV, where a set of studies will be presented.

Methodological Considerations and Procedures

In the field of Artificial Intelligence (AI), it is widely accepted that a system must be capable of generating correct actions through a process of learning, especially autonomous and incremental learning (Pennachin and Goertzel, 2007). To extend this view, this dissertation adds that the process of learning is complemented by a process of development, which progresses in a structured, systematic, and invariant manner. The Hierarchical Stacked Neural Network Model for simulating cognitive development is built out of the MHC and obeys to the general laws that rule hierarchical cognitive development. It is constructed in stacks, where each stack is a computational synonym of a stage.

> Each higher order stack is assumed to change its structure to accommodate the particularities of the higher order stage.

Also, the fundamental characteristics of stacks will be asserted in terms of the fundamental characteristics of stages of development — stages are 1) discrete and 2) cumulative (as mentioned in Chapter II, Section A). Respectively, this means that 1) the transition from one stack to another must be addressed as a discrete process, not continuous, making each stack being partially independent from the immediately lower stack, and that 2) as a higher order stack is formed, every lower order stack must remain available to the network.

In order to implement a system that generates stacks and transits between stacks, it is first necessary to know what a stack is. In the present context, a stack is an independent neural network model that is created to solve a unique and independent set of OHC problems.

> Hence, it is first necessary to find out how it is that each stack represents problem-solving at each OHC. Second, it is necessary to identify what changes from one stack to the next.

Another important aspect of cognitive development refers to the belief that each higher order stage of development is built out of the lower order stage. If this is verified, each higher order stack will be formed out of the higher order stack. Hence, a progression of change will eventually be extracted that can be applied to the generation of a higher stack.

Given the degrees of freedom implicated in searching how a stack represents problem-solving at each OHC, to interpret the changes that undergo from one stack to the next can be very difficult. Moreover, to extract a progression that can be applied to the generation of a higher order stack might be even more difficult. Yet, the following method, similar to the method delineated in Chapter II, Section D, but applied to a computational simulation, aims to initiate such search.

1. Methodological Guidelines

For an AI algorithm, three branches of systems neurosciences must be combined — behavioral psychology, neuroscience, and Artificial Intelligence (Kriegeskorte et al., 2008) — as the present work defends. (Figure 14).

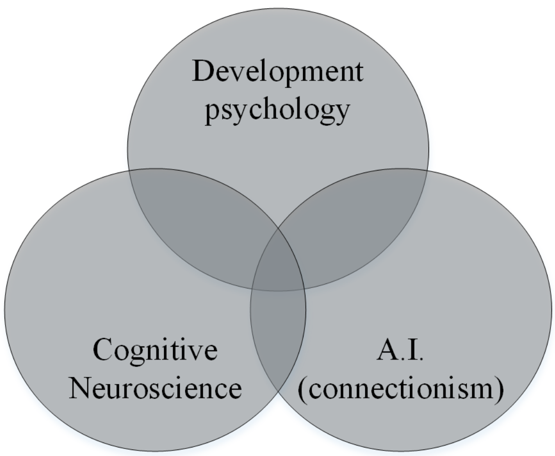

Figure 14 — Overlap between three communicating disciplines

This triangulation is possible because correlation studies between explicit and implicit behavioral observations seem to confirm that the conceptualizations of the mind find a parallelism with neurophysiology. On the other hand, connectionist models were created to mimic some fundamental aspects of neurophysiology. However, the knowledge of brain functioning is itself a theoretical construction, with several conceptualizations having likely explanatory power (Marques-Teixeira, 2013, personal communication). Given that 1) developmental psychology is correlated with brain structure and functioning, and 2) connectionist models were created based on the fundamental principles of brain structure and functioning, one benefits from determining the characteristics of the brain one wants to simulate. Only then, the linkage between developmental psychology and connectionist simulations gains reliability. In fact, one should attempt to bridge levels of explanation in a consistent way across different levels, such that there is isomorphism between levels and areas of description (Johnson, 2011) and reliability in the modelling process (Cassimatis, 2012). It is necessary to increase the commonalities across fields and reach more reliable theories and simulations (Johnson, 2011).

1.1. Conceptualization of the Brain

The brain may be defined ontologically by its anatomical structures, which determine the brain as a “brain”. One may, therefore, define the brain as a ‘structural brain’. Mental states and psychophysiological functions may either be reduced to or identified with the anatomical structures, but they are not considered as ‘constitutive’ for this brain. Anatomical structures, in contrast, are ‘constitutive’ for the brain as a brain and must, therefore, be regarded as both necessary and sufficient conditions for the ontological definition of the brain as a “structural brain”. Since the anatomical structures of the brain can be characterized by physical properties, the “structural brain” may be determined as a physical brain, in an ontological regard. As such it must be distinguished from both “informational brain” and “mental brain” (Marques-Teixeira, 2013, personal communication). Computational Neuroscience is a subfield of AI where simulations of the anatomical brain are conducted. However, it is estimated that a normal adult human “structural brain” is composed of about 86 billion neurons (Herculano-Houzel, 2012) and each of these neurons can have up to 15 thousand connections with other neurons via synapses (Nguyen, 2010). An intuitive assumption is that the accuracy and efficiency with which the brain processes information results not only from the complexity of specialized cell processes, but also from the combined activity of all these cells. Also, aside with the functional complexity of a unit, the information conveyed between units (synapses) is of electrochemical nature, which is mathematically formalized under various parameters — frequency, amplitude, phase, quality and quantity of neurotransmissors, etc. Altogether, this leads to the conclusion that designing a developing architecture by defining the brain as a "structural brain" requires an impeditive exhaustive description of what is (provisionally) known so far, let alone what is not known yet.

The brain may also be defined ontologically by its functions. In this case, functions, or rules, must be regarded as both a necessary and sufficient condition for the ontological definition of the brain as a "functional brain". If the term *functional* refers to physiological functions, the "functional brain" may be determined as a physical brain as well, as physiological functions can be reduced to physical properties. If the term *functional* refers to computational functions, the "functional brain" may be defined as an "informational brain", based on a set of inter-associated rules. If the term *functional* refers to psychological functions, in the sense of mental states, the "functional brain" may be determined as a "mental brain". The "mental brain" is assumed as the coordinated activity of different information processing mechanisms, which, together, create a global pattern of information workflow (Marques-Teixeira, 2013, personal communication). Computational cognition and neuro-informatics are the subfields designated to test models of the cognitive architecture of the brain. The aim is to simulate the interaction of different regions (or cognitive modules), in a way that the model resembles how cognitive functions interact to produce behavior (Anderson et al., 2008; Anderson and Fincham, 2014). This global pattern is usually referred to as a functional or mental state. More recently, some approaches have also used the definition of the term *functional* from the perspective of the Complex Systems Theory (CST).

According to this perspective, processes, functions and rules are invisible to the periphery of the system, but contained in its structure. They are encoded at the level of the components of the system and their internal dependencies (Maturana and Varela, 1928), instantiating that brain functioning is dynamic and mainly implemented in a network (Morita et al., 2016).

In other words, “the response properties of a cortical region are determined by its patterns of connectivity to other regions as well as by their own current activity” (Johnson, 2011).

Differently from the definition of a structural brain, the dynamic functional brain dispenses with the exhaustive definition of all the elements composing the network. What matters is the structure of the model, or system — how many constituents it has and how they are connected with each other — and how the structure encodes the necessary operations. This is in line with the idea that cognitive development is the result of a non-linear dynamic process (Smith, 2005; Smith and Sheya, 2010). This means that “processes and elements of a given stage are more often spoken of as the processes and elements of a dynamic, complex system, and stage change is thought of as the transformation of a system of this kind into another that is more hierarchically complex” (Dawson et al., 2003).

1.1.1. Cognitive development from the perspective of CST

The perspective that cognition is a complex system is actually necessary for understanding that it [the cognitive system] reorganizes itself to solve more difficult problems than those solved before (Mitchell, 1998; Smith, 2005; Spencer et al., 2012). According to the Model of Hierarchical Complexity (MHC), stages of development are seen as equivalents to attractor-states of the system and are represented by increasingly complex structural and functional patterns (Leite et al., 2016). It has also been suggested that each stage of development correlates with a particular neural signature. In other words, it is assumed that a general neural structure exists that evolves in structural complexity (Leite et al., 2016), throughout the so-called stages of development, as interactions with the environment proceed (Mitchell, 1998). A higher-order structure emerges such that the entropy is decreased (Zimmerman and Croker, 2014) by means of self-organization.

The emergence of new more complex functional patterns and the idea of self-organization (Smith, 2005; Spencer et al., 2012) are core characteristics of both the cognitive system in particular, and complex systems in general. They comprise the existence of change within a system. Self-organization means that no single element has causal priority in the explanation of emergence, change, or transition. Transitions are, instead, best assumed to rely on ongoing intrinsic processes of the system as a whole (Mitchell, 1998; Smith, 2005).

The drawback is manifested — processes are, by definition, invisible to the periphery of the system and contained in its structure, hence, difficult to operationalize, observe, and simulate.

Furthermore, self-organization seems to imply that a subsequent stage is built upon the elements and operations of the previous stage (Dawson et al., 2003). Yet, if processes are not seen from the outside, the means by which transitions occur cannot be a priori determined.

1.1.2. Cognitive development in connectionist models

It is here argued that what changes across the progression of problem-solving abilities is the structure of the cognitive and neural system. By structure, we refer to how it is internally arranged such that certain behaviors, or outputs, are produced. By analogy, in connectionist models, structure refers to the number of layers, number of units per layer, and how units are connected to each other within the network. Computational units are distributed information processors, where no single unit has causal priority in the explanation of information processing within the system, as in self-organization. In order to model the *structure* and what *changes* throughout structures underlying stages of development, one needs to identify the structure at each stage, as well as what changes from one stage to the next, such that *structure* and *change* can be coordinated in a model of cognitive development.

> By analogy, if the structure in the human brain changes to accommodate development; the structure of a neural network must change to accommodate development, too.

Throughout development, problems are observed to be systematically solved in an orderly manner, from the easiest ones to the most difficult ones. Then, the first thing to do to ascribe developmental properties to an artificial model is to characterize problem difficulty. The MHC postulates that problems are characterized by an OHC, a unidimensional measure of difficulty. At each stage of development, a unique OHC problems is solved (Commons and Pekker, 2008). According to CST, OHC should be defined in terms of the system's ongoing intrinsic properties, or its structure (Smith, 2005), as much as transitions across orders.

> If processes are contained in the structure, then, evaluating the structure of a connectionist model required for each subset of progressively more complex problems stands as a valid approach for understanding how more complex functional patterns are represented.

The method here devised aims at identifying the connectionist structure that best solves each OHC problem, separately and independently, assuming that each OHC problem is associated with a particular optimal neural networks topology and learning results. The comparison of two structures for solving adjacent and independent OHC problems comprises the representation of transition in problem-solving abilities.

1.2. Biological Plausibility

There is biological plausibility in a priori segregating the problem space into disjoint subsets of OHC problems, such that the structure underlying problem-solving at each OHC is identified. Each stack, or structure, should represent stages of development; stages of development are seen as successive attractor-states of the model. Hence, each stack jointly attracts, or perceives, a specific set of information from the environment, contains rules for problem-solving, and outputs a set of correspondent complexity actions. The emergence of a higher-order stack leads to perceiving the environment as a higher-order set of information, to processing information by higher-order rules, and to outputting a set of higher-order complexity actions, and so on and so forth until the maximum stage of development is reached and the maximum difficulty problems are solved.

By a priori segregating problems by OHC, connectionist models will be created separately, as well. Comparing two adjacent models that were separately created for two subsets of adjacent OHC problems allows for informing about what changes in problem solving abilities from one stage to the next.

Whether successive stacks can be built out of one another will be a matter of experimentation. In fact, one cannot deliberately expect that natural phenomena are mirrored in artificial mathematical models.

One needs to experiment the degree of similarity of both systems in representing the same phenomenon — cognitive development. Until now, the generation of a more complex neural networks structure was done by a sequential addition of network components. Components were added to the existing structure as needed to improve the learning capabilities of the model (Fahlman and Lebiere, 1990). However, as new components are added to the existing network, the generative procedure already implied that each higher complexity structure was built out of the lower complexity structure. Furthermore, the reorganization of existing components has not been studied. Furthermore, if new components are added to the existing structure as they are in current generative architectures, the lower complexity structure is substituted by the higher complexity structure. This contradicts the cognitive capacity to move up and down in problem solving complexity according to the information it receives and does not allow that lower-order structures are protected from the interference of the emergence of higher-order structures.

1.3. Computational Plausibility

Two important works conducted so far within the scope of cognitive development in a connectionist model (Elman, 1993; Norris, 1990) have been mentioned already (Chapter II, Section A). These underline the necessity of building a system in stacks, where each stack solves a less complex portion of the problem than its successor stack. Another work conducted with modular neural networks shows that hierarchical connections between modules is biologically plausible, and computationally, reduces the cost of the entire system and improves results (Mengistu et al., 2016).

Yet, all these works had no general stage theory backing up their computational approach. In other words, they have not a structural criterion for determining what a stack is and what a stack aims to represent.

In the present proposal, the MHC is used to provide a structural basis to the model. Several properties of the MHC provide for its computational representation. The MHC shows that the complexity of human mental activity is incremental in nature and that stages are equally spaced (Commons et al., 2014b), which counterparts a stack approach. Actually, a fundamental idea of stage theory is that cognitive abilities develop throughout a specified and largely invariant sequence that allows no skipping between stages (Commons et al., 2014b; Dawson-Tunik et al., 2005). It shows that the hierarchical configuration of tasks-actions and how it stands for characterizing human development is domain independent (Giri et al., 2014), which shows that the MHC is a structure driven approach and improves the plausibility of the proposed method. Also, it shows that this structure explains human behavior beyond formal stages, which stands for characterizing higher-order cognition, apparently exclusive of human reasoning abilities (Commons et al., 2014a). Finally, the MHC only needs to assume that elements exist, providing an assessment framework suitable for evaluating and comparing human and non-human behavior, including machines (Commons, 2008; Commons and Pekker, 2008).

1.4. The Simulation Context

The simulation context is the Balance Scale Test. It is one of the tests that has been used to study a child's cognitive development (Dandurand and Shultz, 2009; Dawson-Tunik et al., 2010; Siegler and Chen, 2002). Since 1990, it has been repeatedly used as a case-study for computational simulation (McClelland, 1989). Although apparently simple, this tests presents a set of challenging characteristics for simulation, which specifically deal with problem difficulty and difficulty transitions. Although there are many different connectionist simulations for this test, the OHC of its sub-problems has never been manipulated.

The test has been extensively reviewed in Chapter II, Section B. It was early created by Piaget to test children's developmental stage (Inhelder and Piaget, 1958). A few years later, Siegler created an information processing method of the same test so as to standardize behavioral assessment (Klahr and Siegler, 1978; Siegler and Chen, 2002). His version consists of presenting a beam with various weights placed at various distances from the fulcrum, creating different configurations. The MHC measures the difficulty of problems, or configurations, as OHC and ascribes more precision to the behavioral assessment (Dawson-Tunik et al., 2010). The MHC created a pen-and-pencil version of the test, asking children to complete the configurations so that the beam balanced. The simulations conducted in this work will be based on the MHC assessment method, by segregating balance scale configurations by OHC. However, the representation of configurations will remain similar to Siegler's representation, so as to equalize the number and type of output classes per OHC. What matters is that each OHC is represented by a certain operation (Count, Sum, Multiplication, and Distributive Law) (Table 2). Also, current simulations will go up to the Systematic stage because Meta-systematic configurations of the balance scale test present some inconsistencies (not object of the present work). Importantly, this systematic order has been behaviorally tested (Commons et al., 2008; Dawson-Tunik et al., 2010), but has not been tested before in connectionist models.

1.4.1. Data Representing the Balance Scale Test

Data representing all possible configurations of the balance scale test were simulated in two sets, A and B. Order-1 problems correspond to those solved at stage 9 (Concrete) and represent concrete problems; order-2 problems correspond to those solved at stage 10 (Abstract) and represent abstract problems; order-3 problems correspond to those that are only solved at stage 11 (Formal) and represent formal problems (Dawson-Tunik et al., 2010); and order-4 problems correspond to those solved at stage 12 (Systematic) and represent systematic problems, where a coordination between sum and multiplication need to be applied (Table 8).

Table 8 — Operations per OHC problem

Problems	Operation
Order-1 (concrete)	Count how many pegs exist on each side
	Count how many weights exist on each side
Order-2 (abstract)	Sum weight and distance on each side
Order-3 (formal)	Multiply weight by distance on each side
Order-4 (systematic)	Distributive law applied on each side

The first set A is a one-arm beam that contains all configurations, with problems ranging from the concrete to the formal stage. The second set B is a two-arm beam that contains all configurations, with problems ranging from the concrete to the systematic stage (Figure 15).

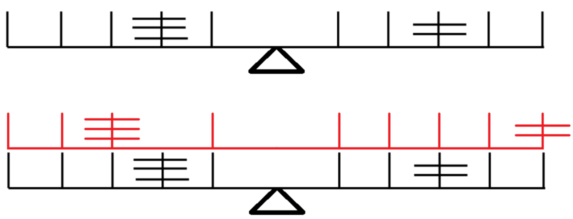

Figure 15 — Representation of a one-arm beam (top beam) and of a two-arm beam (bottom beam)

Balance Scale configurations of set A are represented as 4 integer-element input vectors, where weight and distance values ranged from 1 to

20. Balance Scale configurations of set B are represented as 8 integer-element input vectors (Table 9). Because the possible number of cases for the systematic configurations would be very high, we decided to limit the range of weights and distances from 1 to 5. Outputs were Boolean 3-element vectors that represented one of each three possible classes – fall right [0 0 1], fall left [1 0 0], and balance [0 1 0]. Datasets were segmented so as to allow for controlling for the number of problem dimensions. The systematic-order configuration of the balance scale test increases the number of dimensions for problem solving (composed of two sets of weights per side or consisting in a four-arm beam), and uses the same operations as before — sum and multiplication. A slight change in the operations is present, though. Whereas at the abstract stage only sum was necessary and at the formal stage only multiplication was necessary (Dawson-Tunik et al., 2010; Leite et al., submitted), now both sum and multiplication need to be coordinately applied, by using the distributive law. In particular, the distributive law suggests that the task of evaluating $(a + b) \times c$ is more hierarchically complex than evaluating $(a + b) + c$ or $a \times b \times c$. Hence, the distributive law is of an order of complexity above than sum or multiplication alone and implies a stage transition. Whereas the organization of the actions of addition or multiplication is arbitrary, the organization of the two actions composing the distributive law is non-arbitrary. Therefore, the distributive law is more hierarchically complex than addition or multiplication. Similarly, in the two-part task of first evaluating $(a + b)$ and then evaluating $(c \times d)$ yields the same result as first evaluating $(c \times d)$ and then $(a + b)$. Whenever there is no need of organizing actions in a non-arbitrary way, the MHC says that there is no increase in hierarchical complexity so there is no stage transition (Commons and Pekker, 2008). Otherwise is also true. The dataset segmentation aims at allowing for discriminating the impact of the number of dimensions and the operations conducted.

Table 9 — Representation of Datasets A and B

$$A' = \begin{bmatrix} weight_{right} \\ weight_{left} \\ distance_{right} \\ distance_{left} \end{bmatrix} \qquad B' = \begin{bmatrix} weight\_1_{right} \\ weight\_1_{left} \\ distance\_1_{right} \\ distance\_1_{left} \\ weight\_2_{right} \\ weight\_2_{left} \\ distance\_2_{right} \\ distance\_2_{left} \end{bmatrix}$$

For set A, the problem space was then partitioned into three subsets, each corresponding to each order of complexity (Table 10).

Table 10 — Inputs per order of problem complexity and respective examples for dataset A

Order of configurations	**Operation**	**Input example**	**Expected result**
1	Count how many pegs exist on each side	[2, 2, 12, 16]	[1, 0, 0]
	Count how many weights exist on each side	[2, 12, 11, 11]	[0, 0, 1]
2	Sum weight and distance on each side	[10, 6, 10, 5]	[0, 0, 1]
3	Multiplies weight and distance on each side	[3, 1, 6, 18]	[0, 0, 1]

For set B, the problem space was partitioned into four subsets, each corresponding to an order of hierarchical complexity (Table 11).

Table 11 — Inputs per order of problem complexity and respective examples for dataset B

Order of configurations	Operation	Input example	Expected result
1	Count how many pegs exist on each side	[5, 5, 1, 1, 5, 5, 4, 3]	[1, 0, 0]
	Count how many weights exist on each side	[2, 4, 5, 5, 4, 1, 5, 5]	[1, 0, 0]
2	Sum weight and distance on each side	[3, 5, 5, 3, 5, 4, 1, 1]	[0, 0, 1]
3	Multiply weight by distance on each side	[5, 3, 1, 1, 2, 4, 2, 2]	[0, 0, 1]
4	Distributive law applied on each side	[5, 3, 2, 5, 2, 4, 5, 2]	[1, 0, 0]

Data was, then, partitioned into training (70%), validation (15%), and test (15%) per each order of complexity problem. The number of cases per class differed per experiment, so each experiment in the following chapter will include a reference to that.

1.5. Neural Networks Models

There is a difference between applying this method to the field of developmental cognitive neuroscience and to the field of AI. The difference is that while looking into the brain allows to see how stages are (already) represented, in a connectionist model, the structure underlying each stage needs to be found. To find out how it is that each stack represents problem-solving at each sequential OHC can be extremely difficult for two main reasons. First, the number of hyper-parameters that influence learning in a neural network model in combination with the characteristics of the inputs it receives is huge. Second, there is a common risk associated with simulating cognitive phenomena with ANN. If a network has a sufficient number of units connected with each other, the information will be sufficiently distributed throughout the network. This will result in a system where inputs will always be accurately transformed into outputs.

The risk is that the researcher is falsely led to believe that cognitive procedures have been simulated with biological reliability, while the artificial system has only been able to create a mathematical mapping between inputs and desired outputs (Cassimatis, 2012).

The challenge, then, is to initiate the search of the minimal complexity connectionist structure that solves cognitive problems of each OHC, separately and independently, such that it [the connectionist structure] is more likely to inform about the possible structure of information processing of the nervous system per stage of development.

Networks per each OHC problem of the balance scale test were created. For order-1 problems, two networks were created, one for solving weight problems, and another for solving distance problems. This is based on the notion that for order-1 problems (stage 9), the cognitive structure either solves weight problems or distance problems separately, indicating that children shift their attention to one or the other dimension alone.

1.5.1. Networks Structure

Several neural network models were trained to solve each OHC problems separately and independently. These networks were created using the neural network toolbox available in MatLab®, R2016b. All hyper-parameters and stopping criteria were defined heuristically and kept constant throughout all experiments, with three exceptions. Exceptions were the number of units per layer, the number of layers, and the connectivity pattern among units (Figure 16).

In order to find the minimal complexity structure, units per layer were added sequentially, one-by-one. Networks were composed by either an internal layer (perceptron networks) or two layers (hidden-layer networks) plus the output layer. For hidden-layer networks, for each number of units in the internal layer (first layer), hidden units were sequentially added in the hidden layer (second layer). Units were added until one of two stopping criteria was reached — either if the current network reached a learning accuracy of 100% or if the number of units had achieved the maximum. The upper bound of the number of units N_{hu} was set equal to 20, according to (Equation 1):

$$N_{hu} = \frac{N_s}{p \times (N_i + N_o)} \qquad \text{Equation 1}$$

where N_s is the number of training samples, N_i is the number of input units, N_o is the number of output units, and p is an arbitrary scalling factor, usually set between 2 and 10, here set equal to 7. This resulted in a maximum of 23 units for order-2 problems and a total of 35 units for order-3 problems. However, during training, it was determined that adding more than 20 units per layer, in both order problems, was not informative. The maximum number of units was, then, based on the above equation but adapted to the current case. In the case of perceptron networks, 20 possibilities were tested per OHC; in the case of hidden-layer networks, 400 possibilities were tested (20 units in the internal layer × 20 hidden units). Each possibility was repeated over 20 trials to compensate for the random generation of initial weights.

In regards to the connectivity pattern among units, in the first and second studies, all networks were feedforward fully-connected architectures. In the third experiment, five different connectivity patterns were tested, as specified in the respective subsection of the following chapter. Per each connectivity pattern, the procedure of addition of units and layers remained, as well as the stopping criteria. In this experiment, only models for order-3 (Formal) and order-4 problems (Systematic) were trained, due to the results of previous experiments. Models for order-3 problems were trained using perceptron and hidden-layer networks. If stopping criteria did not apply, a total of 20 perceptron networks (1 to 20 units in the internal layer) and 400 hidden-layer networks would be tested (20 units in the internal layer × 20 units in the hidden layer) per connectivity pattern (5 types). In total, 2020 networks were trained. For systematic problems, given that only hidden-layer networks were trained, a total of 2000 networks were trained.

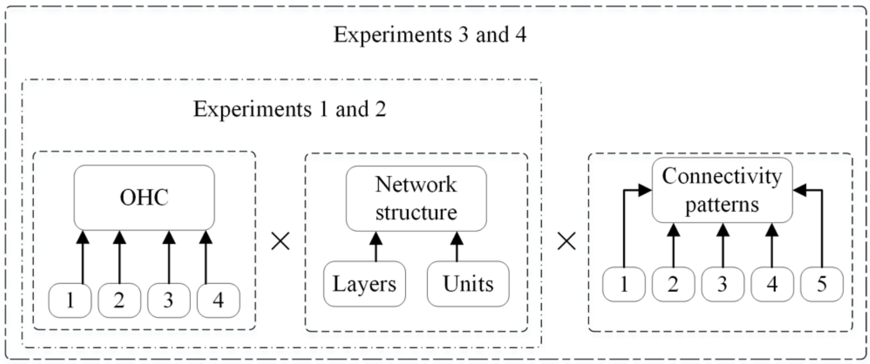

Figure 16 — Figure scheme of experiments

1.5.2. Networks Learning

Activation Functions: The sigmoid activation function was set to the hidden units; the normalized exponential function (softmax) was set to the output units, as the present learning environment consists of a classification task.

Initialization weights: In experiments 1, 2, and 3, weights were initialized according to the Nguyen-Widrow initialization algorithm, which chooses values in order to distribute the active region of each neuron in the layer approximately evenly across the layer's input space. The values differ each time the network is initialized. In experiment 4, the weights of the lower order network, after training, were the initial weights of the higher order network.

Learning Algorithm: The gradient descent algorithm with adaptive learning rate (*lr*) was chosen, with initial *lr = 0.01*. The learning rate changes as the network outputs a success or a failure during training. At each epoch e, if $\frac{\text{error}(e)}{\text{error}(e-1)} > 1.04$, then, the $lr_{(e+1)} = 0.7\ (lr_{(e)})$ and weights and biases of epoch e are discarded. Otherwise, $lr_{(e+1)} = 1.05\ (lr_{(e)})$ and the weights and biases of epoch e are kept.

Cost function: The cross-entropy loss function (CE), calculated as (Equation 2)

$$CE = -t \times \log(y) \quad \text{Equation 2}$$

Stopping criteria per network: A maximum of 10 validation steps or a minimum performance of 0.001 were set. As soon as one of these were reached, the network would stop training.

1.5.3. Networks Performance

All trained networks were evaluated based on mean total accuracy (A_t), total number of connections (N_c), and an inverse measure of efficiency (EF). A_t is the mean total accuracy of networks averaged across each set of 20 trials. EF was calculated to identify the optimized point between increase in accuracy and increase in computational cost as units were added (Equation 3). Here, EF_{hu} represents the inverse efficiency value per number of units, N_c represents the number of total connections of the network, and $e^{1.5}$ represents the mean error generated by the network potentiated to an arbitrarily defined parameter of 1.5. This parameter overweighs an increase in accuracy against an increase in cost such that successes are reinforced. After calculating the *EF* value, the difference in *EF* ($Diff_{EF}$) was calculated (Equation 4). Negative values of $Diff_{EF}$ correspond to local minima of the *EF* function, which correspond to increases in Efficiency and allowed for selecting the best network per order of complexity problem.

$$EF_{hu} = \frac{1}{N_c \times e^{1.5}} \quad \text{Equation 3}$$

$$Diff_{EF} = EF_{hu} - EF_{hu-1} \quad \text{Equation 4}$$

After selection, network structures of adjacent orders of complexity were compared in terms of changes in structure and accuracy. The fourth study used the selected networks across experiments 2 and 3 to extract a progression of change throughout network structures of adjacent OHC. For that, a difference in the hyper-parameters was included, regarding the initialization weights, as mentioned above.

2. What Next

The present section exposed a method that will, hopefully, allow for answering the research question: “When a developmental transition occurs, the system changes from what to what, and how?”. For that, two steps need to be accomplished:

1) to determine the structure of each stack per OHC and 2) to compare the structure of adjacent stacks, such that, eventually, a progression of change (or connectionist development) can be determined.

CHAPTER IV

EXPERIMENTAL STUDIES FOR SIMULATING COGNITIVE DEVELOPMENT IN A CONNECTIONIST MODEL

The methodological considerations and procedures delineated in the previous Chapter III are transversal to all the four studies that are contained in the present Chapter IV. The present chapter begins with a preliminary section, which highlights the main findings. The section containing the four studies follows. The first and second studies, respectively, respond to the third objective "Identification of the factors underlying the representation of stages of development". They investigate the relative influence of the number of problem dimensions and the required operations for successful problem-solving at each stack. The fourth objective "Modelling structural changes across connectionist stacks" is responded by all the four studies.

The first two, already mentioned, are two-folded, then. They also investigate how problem-solving of different OHC are represented in segregated connectionist stacks, when the number of units and layers is allowed to vary. The third study investigates how OHC are represented if the connectivity pattern among units also varies. Finally, the fourth study builds on the results of previous studies and shows that the previously found connectionist structures can emerge out of one another, forming a Hierarchical Stacked Neural Networks that grows in structural complexity, simulating cognitive development.

Highlights of Chapter IV

- Connectionist structures per stack are able to represent the OHC by varying the number of units per layer, the number of layers, and the connectivity pattern among units
- Connectionist structures per stack increase in complexity (concerning the number of units and layers) in a non-linear way
- If balance scale problems are segregated by OHC, formal problems are solved with 100% accuracy
- There are two types of structural transitions
- Memory-based occur when the number of problem dimensions increases, from the concrete to the abstract stage
- Operationally-based occur when the complexity of the required operation increases, from the abstract to formal and from formal to systematic stage, exclusive of higher-order cognition
- In terms of developmental progression, it was possible to identify that a higher-order structure included the components of a lower-order structure, even with networks being trained separately and completely independently
- Different departing structures can be the starting point for higher-order structures, if adjacent structures share the connectivity pattern
- The densest connectivity pattern for networks with input connectivity showed the best improvements in performance, the highest resistance to learning rate modifications, and the best plausibility in the structural growth process
- There was a slight tendency for networks with different number of units and different connectivity patterns converging to the same number of connections for formal and systematic problems
- This suggests that, instead of Hierarchical Integration of lower order actions, or outputs, one might talk about "hierarchical structural integration", where "the higher-order structure is formed out of the lower-order structure"

Connectionist Models Capturing Stages of Development and Stage Transitions

In order to maximize the similitude between human cognition and artificial models, one should adopt a longitudinal perspective-taking and simulate the development of cognitive abilities. The idea is to build an algorithm that not only *learns*, but one that also *develops*. As connectionist models are widely used to learn many complex problems, they started to be used to solve a more intriguing problem — the problem of solving a task *developmentally*. This implies learning simpler problems first and more complex ones later, sequentially (McClelland, 1995, 1989; Norris, 1990; Shultz and Schmidt, 1991), which has not been an easy task. Coordinating learning and development in a connectionist model is the starting point to ascribe hierarchical flexibility and stability to the system.

The difficulty of segregating learning and development in connectionist models is intrinsic to the nature of learning in these models. They are global compositions of units and weighted connections linking those units, or single-corpus models. Previous work has shown that the model needs to change its structure as more difficult problems are presented. An artificial model of this kind is called a generative architecture and has been shown to best approximate a stage-like structure and stage-like performance in cognitive developmental problems (Dandurand and Shultz, 2009; Elman, 1993; Fahlman and Lebiere, 1990; Leite et al., submitted; Norris, 1990; Shultz et al., 1994). Yet, as new components are added, all weights (the old and the new) change to form a new global structure. Furthermore, the properties of problems that triggered the recruitment of more candidate components are not a priori set. This results in older structures no longer available. For this reason, even existing generative models are single-corpus models, lacking the capability of *hierarchical* flexibility and stability, or hierarchical adaptation.

A developing connectionist model should, thus, transform classical connectionist models from a single-corpus of knowledge to a step-wise corpus that *develops* by stacks.

This fundamental change implies that each connectionist structure constitutes a stack and is only formed after the immediately lower order one has been formed, trained, and kept in memory. The algorithm needs not only to *learn* the optimal weight combination per stack, but also to *develop* from one stack to the next. The simulation of stage transition concerns the ending of training of a lower-order stack and the formation and initialization of training of a higher-order stack.

Developmental studies are consistent in showing the existence of periods of consolidation (stages of performance) and transitional periods (stage transitions). Stages are characterized by stable and largely homogeneous performance, whereas stage transitions are characterized by U-shaped performance (Dawson-Tunik et al., 2005). Stacks are, then, the computational synonym of stage of development and stack transition is the computational synonym of stage transition.

Two important questions rise. 1) What homogenizes performance within stage and how can stages be simulated in an artificial model? The Model of Hierarchical Complexity (MHC) is a general stage theory that attributes to problems a unidimensional, abstract, linear and equally-spaced order of hierarchical complexity (OHC). The OHC characterizes stages of development, predicting cognitive capacity with high accuracy (Giri et al., 2014). Hence, the MHC introduced an objective and accurate framework for explaining homogeneity of performance at each stage. However, it is not known yet how a connectionist model would represent OHC. Previous work points towards the interaction between operations and input dimensions, but this needs to be clarified. The second question is 2) how are stage transitions processed and how can they be simulated in an artificial model? Unfortunately, the mechanism underlying stage transitions is not clarified. Hierarchical Integration (HI) has been proposed at a behavioral level, but it finds little to no evidence in brain-based experiments and has been invalidated in previous work using connectionist models. Nonetheless, stage transitions have been object of attention by artificial learning modelers, who strive for building algorithms that learn to solve tasks nearly as efficiently and flexibly as humans do.

1. Summary of Previous Simulations

In developmental and evolutionary cross-species studies (Commons et al., 2014c; Leite, 2016), it has been identified that only humans develop cognitively until they perform formal operations to solve problems. Formal operations no longer require the concrete experience of the problem. They refer to the ability of operating with abstract and theoretical concepts and using logic to creatively find new solutions. Pioneering theories concerning developmental cognition consider that the acquisition of formal operations is the highest point in problem-solving abilities, opening up the way for higher-order cognition (Inhelder and Piaget, 1958). It has been soon proposed that cognitive models of artificial learning should capture a similar progression in problem solving abilities as seen in living organisms (McClelland, 1989), until higher-order cognition is achieved. This idea was pioneering in setting forth the overlap between developmental psychology and computational cognition.

The balance scale test is commonly used as a context for reproducing stages of cognitive development in connectionist models (Dandurand and Shultz, 2009; Dawson and Zimmerman, 2003; Leite et al., submitted; Leite and Rodrigues, 2018; McClelland, 1989; Reyes et al., 1997; Schapiro and McClelland, 2009; Shultz et al., 1994; Shultz and Cohen, 2004; Shultz and Schmidt, 1991; Zimmerman, 1999; Zimmerman and Croker, 2014). The goal of these simulations is to ascribe to the connectionist model the ability to solve problems through the same sequence as is observed in children — from the simpler to the most difficult problems.

All connectionist simulations of the balance scale test are based on Siegler's approach and refer to the acquisition of the fourth Piagetian stage, or eleventh MHC's stage. All models are fed with all possible configurations of the balance scale test. Their performance is evaluated in terms of which problems are solved before which and how accurately. All these simulations depart from a pre-defined connectionist structure and learning algorithms that ideally represent the sequence of Siegler's rules and reproduce the sequence of problem-solving capabilities. Pre-determined does not mean fixed, as some work has been conducted with generative topologies. Simulations based on Siegler's assessment method have limitations, which mainly relate to the controversy of defining difficulty of problems (homogeneity within stage) and difficulty transitions (stage transitions).

These limitations poorly answered the first question posed above — what homogenizes performance at a given stage? — hence, they poorly allowed for building developmental models of human cognition.

For instance, Siegler postulates that the turning point in difficulty concerns the transition from solving problems where only weight or only distance on each side vary to solving problems where both dimensions vary (Siegler and Chen, 2002), referring to the first transition. Yet, simulations show that the most difficult transition refers to the ability of using the multiplication operation to solve torque problems (Reyes et al., 1997; Shultz et al., 1994; Shultz and Schmidt, 1991), referring to the second transition. Namely, intuitive networks, identified as those which learn only from examples, perform well on all problems requiring rules I (weight), II (distance), and III (large torque conflict problems), but fail to represent rule IV (small torque conflict problems that require multiplication). This is consistent with the acquisition of formal operations being the turning point to higher-order cognition, and raises doubts as to whether the theory accurately captured the underlying developmental mechanisms.

First attempts to simulate human performance on the balance scale test assumed the perspective that development is continuous (McClelland, 1995, 1989). Discrete stage transitions were not taken into account until one specific work showed that only a generative architecture could deal with the most difficult subset of problems (Shultz and Schmidt, 1991). From then on, cascade-correlation neural networks (Shultz et al., 1994; Shultz and Schmidt, 1991), and variants of cascade-correlation that attempt to simulate the achievement of higher-order rule IV (Dandurand and Shultz, 2009; Reyes et al., 1997) have been used. Generative architectures are those neural network structures to which units are added as problems become more demanding. An indicator of the difficulty of problems is performance dropping. The addition of units increases the distributive learning potential of the network, decreasing its error again (Fahlman and Lebiere, 1990). With these architectures, although rule IV problems could be accurately solved (multiplication problems), the model jumped over problems requiring rules I and II (Shultz et al., 1995). This is because higher-order structures take over lower-order structures. The solution encountered that best solves this subset of more difficult problems has been to inject a torque function, by using a Function Based Cascade Correlation architecture (Dandurand and Shultz, 2009), or to inject a new assimilation function — weighted product instead of weighted sum (Reyes et al., 1997).

These studies (Dandurand and Shultz, 2009; Reyes et al., 1997) already suggest that the structure underlying difficulty transitions is a generative one and that the structure for solving the more difficult problems needs to change more abruptly. However, the growth in difficulty and problem-solving abilities has yet to be explained by any of the previous models, as no previous work justifies why intuitive networks perfectly solve problems requiring rules I, II, and III, but fail to solve problems requiring rule IV. Also, it has yet to be clearly shown what it is that changes structurally in both transitions.

> None of existing simulations based on Siegler's approach could yet clearly show what is at stake in developmental transitions, specifically in what concerns the transition to formal reasoning (Dandurand & Shultz, 2009; Reyes et al., 1997).

In terms of transition across stages, the MHC and several other authors have proposed that these involve transformations such as HI (Case, 1987; Demetriou and Valanides, 1998; Fischer, 1987), a perspective that prevailed until nowadays (Commons and Pekker, 2008). HI is primarily a Piagetian concept, which postulates that outputs generated at a lower-stage of performance become object of outputs to be generated at the immediately next stage. Previous simulations, exposed in Chapter II, Section C, tested the mechanism of HI (Leite & Rodrigues, 2018). Results showed that it does not apply to the formation of higher-order outputs. Specifically, if the model was only fed with lower-order outputs, it would not learn to generate higher-order ones. If the model was fed with both lower-order outputs and a different set of inputs, such as the original percepts composing the problem, the model would never value the information conveyed by lower-order outputs for the generation of higher-order outputs.

The generation of increasingly complex outputs during development is here assumed to be the result of the emergence of a new organizational and functional pattern. The term "new pattern" applies to behavior, to the neural arrangement, and, presumably, to the structure of connectionist stacks. Thus, the growth in stage of development always corresponds to the generation of a higher-order stack, which implies stack transition. Development enables the system to find more complex states of equilibrium, which account for solving increasingly complex problems, producing increasingly complex operations, and generating increasingly complex outputs. Differently, learning refers to the fine-tuning of the active stack. Hence, learning occurs within stack, or within stage of development. Learning at each stack refers to "consolidate each stage so that it is protected from interference caused by learning in the following stages" (Norris, 1990).

2. Present Simulations

Stage transitions, or stack transitions, are the object of study of the following set of four experiments, using connectionist models. This set of studies aims at understanding how a connectionist structure grows in structural complexity to solve increasingly complex problems (Commons, 2008), as occurs throughout development. This set of four experiments responds to the second and third objectives of the present dissertation. Namely, the identification of important factors underlying the representation of stages of development in a connectionist stack (Experiments 1 and 2), and the modelling of structural changes across connectionist stacks (Experiments 1, 2, 3 and 4).

Networks performance is depicted in graphs, when relevant. All graphs depicting this information are similar. Those depicting the performance of perceptron networks show the mean total accuracy (*At*) and standard deviations, total number of connections (*Nc*) and efficiency (*EF*), as units were added in the internal layer, from 1 to 20. Y-coordinate in *At* values represents the mean *At* calculated over all 20 repetitions. *EF* was calculated as a function of the represented *At* value. The graphs depicting the performance of hidden-layer networks show the mean total accuracy (*At*) and standard deviations, total number of connections (*Nc*) and efficiency (*EF*), as units were added in the internal layer, too. Yet, the Y-coordinate in *At* values represents the highest mean *At* found across units added in the hidden layer. The number of units in the hidden layer correspondent to the highest value of *At* is labelled at each point of *At*. EF values were calculated as a function of the represented *At* value.

EXPERIMENT 1

In the balance scale test, the first transition implies an expansion of the range of dimensions considered for problem solving. Children go from considering weight or distance alone, to considering weight and distance together. Differently, the second transition implies that children learn to coordinate weight and distance such that small torque differences between the sides of the scale are solved (Dandurand and Shultz, 2009; Dawson-Tunik et al., 2010; Shultz and Schmidt, 1991; Siegler and Chen, 2002). This second transition implies not an increase in the range of dimensions, but an update of the operations that represent the coordination of weight and distance. The fact that this second transition is different from the first one suggests that simulations must not only describe the performances at each stage, but must also be open to the possibility that transitions from one level of difficulty may be different from those at another level. In other words, if transitions in difficulty rely on either the number of dimensions or on the operation that coordinate the input dimensions, how exactly is difficulty represented in connectionist models?

> Is the increase in hierarchical complexity of problems a matter of the number of problem dimensions or a matter of the operations used to solve it?

This experiment has a two-folded objective. It aims to determine 1) the influence of the number of problem dimensions and the operations conducted for representing preformal and formal stages of development in connectionist stacks, and 2) the structural changes from preformal to formal stages, based on the number of computational units and layers. The structure of a connectionist model for each OHC problem will be analyzed and models of adjacent OHC problems will be compared.

Experiment 1 was conducted using dataset A, containing configurations from concrete to formal problems. From a total of 160,000 possible configurations, 15,600 were order-1, 129,860 were order-2, and the remaining 14,540 were order-3. In order-2 and order-3 problems, the number of cases per class was balanced, which reduced the subsets to 1140 and 1740 cases, respectively.

Results of Experiment 1

Networks of order-1 and order-2 problems required the same topology for learning with 100% accuracy — a minimum complexity structure of 1 computational unit and 12 connections. Mathematically speaking, both order-1 and order-2 problems are linearly separable. Hence, a single-unit perceptron network was sufficient. Only when contrasting order-2 and order-3 problems did the structure of the networks changed significantly. This is congruent with the fact that order-3 problems are not linearly separable. Also, from order-2 to order-3 problems, there is a change in the operator required to solve the problem, referring to the emergence of formal reasoning, while the number of dimensions remains the same.

For order-1 and order-2 problems, once the topology was similar, changes in the weight matrices of all three networks (weight problems, distance problems, and conflict large torque problems) were inspected (Table 12 and Table 13). Table 12 represents the weights that connect each of the four input units to the single unit in the internal layer of order-1 and order-2 networks. Table 13 represents the weights that connect this single unit to the 3 units of the output layer in order-1 and order-2 networks.

Table 12 — Weight matrix between input and internal layers

	$\text{weight}_{\text{righ}}$	$\text{weight}_{\text{left}}$	$\text{distance}_{\text{right}}$	$\text{distance}_{\text{left}}$
Order-1 distance network	0.000	-0.001	6.309	-6.308
Order-1 weight network	5.722	-5.776	0.005	-0.005
Order-2 network	4.069	-4.062	4.071	-4.074

Table 13 — Weight matrix between internal and output layers

	output unit 1	**output unit 2**	**output unit 3**
Order-1 distance network	9.458	0.239	-9.026
Order-1 weight network	8.538	0.211	-8.078
Order-2 network	8.646	0.269	-8.206

Table 12 clearly shows that networks for order-1 problems represent a shift from one dimension to the other. Either distance or weight have non-zero weights, which is quite obvious due to the structure of inputs (either weight or distance alone varied). In the network for order-2 problems, both dimensions — weight and distance — become coordinated. This coordination is represented by a superimposition of the weight matrices of the networks for both order-1 problems. Specifically, the right side is represented by positive weights and the left side is represented by negative weights of the same magnitude. This is interesting given that order-1 networks were trained separately, with random initial weights. In what concerns the weights between the internal layer and the output layer (Table 13), no significant differences were found.

In what refers to order-3 problems, the required structure changed significantly. Table 14 identifies the preferential topology of each order-3 networks, both perceptron and hidden-layer networks, based on negative values of $Diff_{EF}$. More than 3 negative values were found. In the table, only those of the same order of magnitude are described.

Table 14 — preferential topology of each set of networks for order-3 problems

Network type	Diff EF	Units (internal layer)	Units (hidden layer)	Connections	Mean Accuracy (%) (∓ sd)	Max accuracy (%)
Perceptron networks	-5.95	12	—	99	73,60 (± 2.43)	77,18
	-1,47	3	—	27	59,85 (± 7.86)	67,15
	-0,04	20	—	163	74,93 (± 2.79)	79,37
Hidden-layer network	-418,12	14	7	199	67,85 (∓4,16)	81,32
	-208,51	10	9	179	67,13 (∓3,67)	77,41
	-198,36	19	11	351	68,90 (∓10,70)	85,57

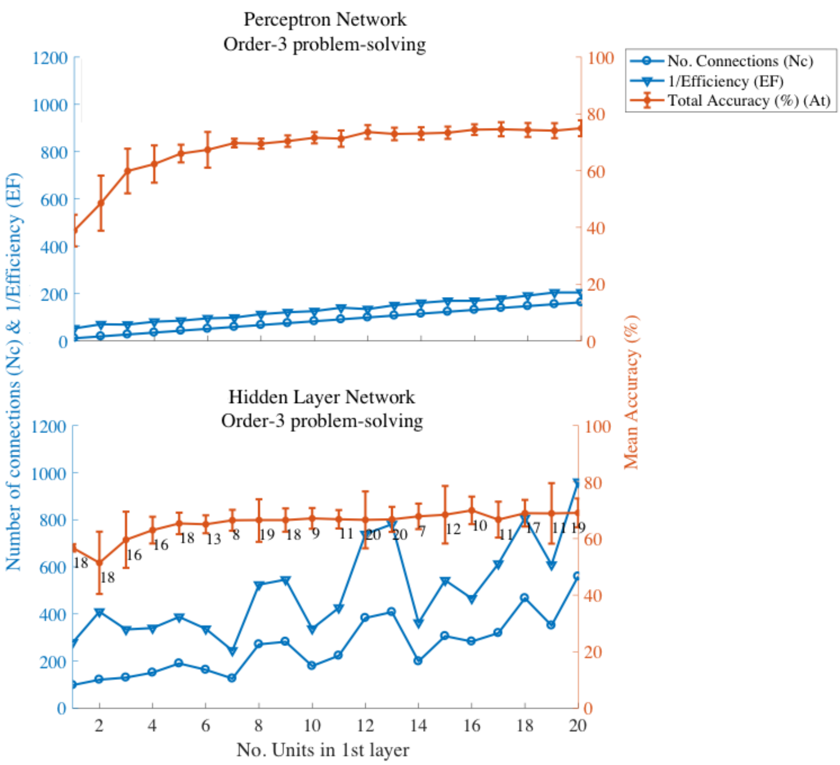

Figure 17 — Topology and performance of perceptron networks and hidden-layer networks in total accuracy (*At*), total number of connections (*Nc*) and Efficiency (*EF*) for order-3 problems

From the graphs (Figure 17), it is possible to observe that hidden-layer networks have higher variability due to the higher number of connections among units. Also, the difference between mean At and highest At is greater for hidden-layer networks than for perceptron networks (Table 14). Nonetheless, hidden-layer networks have a higher potential to learn the solution more accurately. It is important to state that a higher variability, in the present case, is not due to the ratio between number of training cases and total number of connections, as these ratios are approximately similar for perceptron networks and hidden layer network.

Discussion of Experiment 1

The most informative result is the confirmation that the two transitions comprising the balance scale test are of a different nature and that this difference is captured by the structure of networks.

In the first transition, from order-1 to order-2 problem-solving, although the number of units and connections do not change, the weight matrices did change in a significant manner. The weight matrices of both networks for order-1 problems were superimposed to form the weight matrix of the network for order-2 problems. This is interpreted as a computational synonym of accommodation of previous knowledge. In a coarse-grained interpretation, this suggests a transition in difficulty perhaps based on a memory expansion. According to neuroscience developmental findings (Qin et al., 2014), as children learn to operate with a specific set of information, knowledge about specific problem-solving becomes stored in hippocampal memory. This results in a faster and less consuming retrieval of information, which allows new knowledge to be acquired (Zimmerman and Croker, 2014). A possibility about the interpretation of the first transition is that it captures predominantly on such memory-based mechanisms, allowing for order-1 knowledge to be chunked to form order-2 knowledge (Sweller, 2004), according to the present simulation context. In terms of the operation conducted, from counting to summing, one can deduce that summing is a matter of counting with more elements. In fact, counting is represented by the same operator "+".

In the second transition, the number of dimensions remains the same, and the operator changes. While order-2 problems require the use of sum, order-3 problems require the use of multiplication. For this reason, and based on previous studies (Dandurand and Shultz, 2009; Reyes et al., 1997), a complexity jump in information processing was identified. This jump consists in a shift from a perceptually-based to a formally-based consideration of the problem. It is easy for children to coordinate weight and distance when they are manipulating objects, but it is much harder to put together cause and effect for two variables (Zimmerman and Croker, 2014), which characterizes formal reasoning (Dawson-Tunik et al., 2010). That was seen in the results. Neural networks for solving order-3 problems required a more complex structure to encode the multiplication operation. In line with previous studies (Richardson et al., 2018), adding a hidden layer showed a greater potential to encode order-3 problems than increasing the number of units in a perceptron network, although other properties and functions should be tested. These results indicate that from order-2 to order-3 problems there is a structurally deepest transition in difficulty, suggesting that orders of hierarchical complexity are not equally spaced (Commons et al., 2014b), at least if represented in a connectionist model of the present type.

EXPERIMENT 2

Experiment 2 was conducted using the dataset B, comprising problems from concrete to systematic.

Given that dataset B consists of a two-arm beam, the impact of the number of problem dimensions used to characterize the same set of problems will be evaluated.

Furthermore, structural changes from the concrete to the systematic problem-solving will also be evaluated. According to the MHC, it was hypothesized that the connectionist structure for solving systematic problems is more complex than that for solving formal problems, due to the distributive law, and that the transition from formal to systematic is an operationally-based transition. Yet, according to the distributive property of connectionist models, one can, as well, expect that the new coordination rule is a matter of information distribution conducted within the model, once the arithmetic operations for the systematic stage are the same (sum and multiplication). The present experiment, then, aims at 1) determining the influence of the number of problem dimensions at all OHC, given that all OHC problems are represented in a two-arm beam, and at 2) determining the nature of the transition from formal to postformal stage.

The number of cases per class was balanced, using the number of "balancing cases" as a reference. At last, there were 1011 cases of order-1 configurations, 5790 of order-2, 1128 of order-3, and 658 of order-4. Data was, then, partitioned into training (70%), validation (15%), and test (15%) for each order of complexity problem.

Results of Experiment 2

As in the previous experiment, networks for order-1 and order-2 problems learned with maximum accuracy (100%) with the minimal structure of one unit and a total of 12 and 16 active connections, respectively. Namely, order-1 networks for weight problems did not use the four connections responsible for encoding distance, and otherwise.

For order-1 problems, the weight matrices of both weight and distance problems networks clearly show the shifting of attention between both dimensions (Table 15). In networks for order-2 complexity problems, the weight matrix between the input and internal layer illustrates that order-2 problem-solving coordinates both weight and distance inputs, by superimposing both weight matrices of networks for order-1 problems. Learning epochs and time do not allow for discriminating the performance of the models (Table 16).

Table 15 — Weight matrices between input and internal layers

	$W_r(1)$	$W_l(1)$	$D_r(1)$	$D_l(1)$	$W_r(2)$	$W_l(2)$	$D_r(2)$	$D_l(2)$
Order-1 (distance prob.)	0,00	0,00	-3,61	3,59	0,00	0,00	-3,60	3,58
Order-1 (weight prob.)	-3,13	3,13	-0,01	0,00	-3,12	3,12	0,00	0,01
Order-2 network	-3,40	3,40	-3,40	3,40	-3,40	3,40	-3,40	3,40

Table 16 — Performance parameters of order-1 and order-2 networks

	Accuracy	# Epochs	Best epoch	Time (s)
Order-1 (distance prob.)	100%	221	211	0,55
Order-1 (weight prob.)	100%	175	175	0,43
Order-2 network	100%	250	244	0,83

Also, similarly to the previous experiment, the performance of networks for solving order-3 and order-4 problems required a more complex structure. Their performance is depicted in Figure 18 and Figure 19, respectively. Following the figures, Table 17 and Table 19 identify the preferential structure of each set of selected networks (perceptron and hidden-layer) for each OHC problem, based on negative values of $Diff_{EF}$. Table 18 and Table 20 describe the performance of each selected network, also for each OHC problem. In each pair of tables (for order-3 and order-4 problems), the selected networks are numbered to facilitate comparison across tables.

Networks for solving order-3 formal problems, composed of 12 units and 147 connections, learned the problem perfectly, performing with 100% on the test set. In Figure 18 (top graph), it can be observed that networks become increasingly efficient after the addition of 5 units, reaching maximum performance with 12 units, showing that the increase in units is highly favoring the increase in accuracy.

Hidden-layer networks were also trained on order-3 formal problems to evaluate whether the addition of another layer of computations would maintain accuracy and decrease the number of total units and total connections (Figure 18, bottom graph). Yet, that was not verified (Table 17). Furthermore, efficiency in hidden-layer networks is highly non-linear due to the increased number of connections.

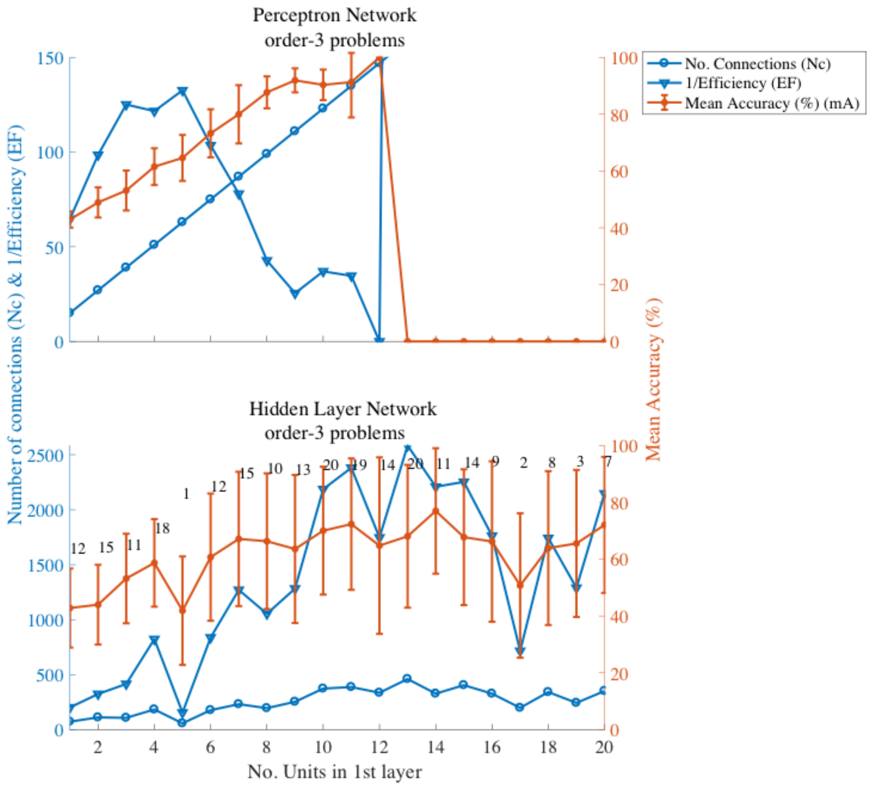

Figure 18 — Topology and performance of perceptron networks and hidden-layer networks in total accuracy (*At*), total number of connections (*Nc*) and Efficiency (*EF*) for order-3 problems

Table 17 — Preferential topology of network for order-3 problems

Network type	**Diff EF**	**Units (internal layer)**	**Units (hidden layer)**	**Connections**
Perceptron network	—	12	—	147
Hidden-layer network	-1046,78	17	2	198

Table 18 — Performance of best network for order-3 problems

Network type	**Mean Accuracy (∓ sd)**	**Max Accuracy**	**Epochs**	**Best Epoch**
Perceptron network	—	100%	519	509
Hidden-layer network	50,73% (∓25,42)	99,20%	738	727

Networks for solving order-4 problems required a more complex structure than networks for order-3 problems. Their maximum accuracy was lower. This was true for both perceptron networks and for hidden-layer networks (Table 19).

Interestingly, for perceptron networks of both orders, different performance patterns were found. For order-4 problems, the addition of units in perceptron networks allowed only for a small and subtle increase in learning accuracy, which resulted in an approximately linear decrease in networks' efficiency (Figure 19, top graph). Furthermore, only after the addition of 15 units, did the perceptron networks for order-4 problems learned with an accuracy above 86% (but never surpassing 87,8%). In sum, while a perceptron network for order-3 problems benefitted from the addition of units and reached total accuracy (100%) with a minimum of 12 units, perceptron networks for order-4 problems seemed to face an increase in the computational cost as units are added, requiring a minimum of 16 units to reach lower accuracy (> 86%).

Then, hidden-layer networks were trained to solve order-4 systematic problems and to evaluate whether the addition of another layer would have a positive effect on the network's performance. It was found that networks with 10 units, or more, in the internal layer and a variable number of units in the hidden layer achieved an accuracy bounded by 85% and 87,7%, which was similar to perceptron networks with a minimum of 16 units (Table 19). Furthermore, in hidden-layer networks for order-4 problems, accuracy was highly variable across trials (Table 20), as well as their efficiency as units were added. This was due to the increased number of connections.

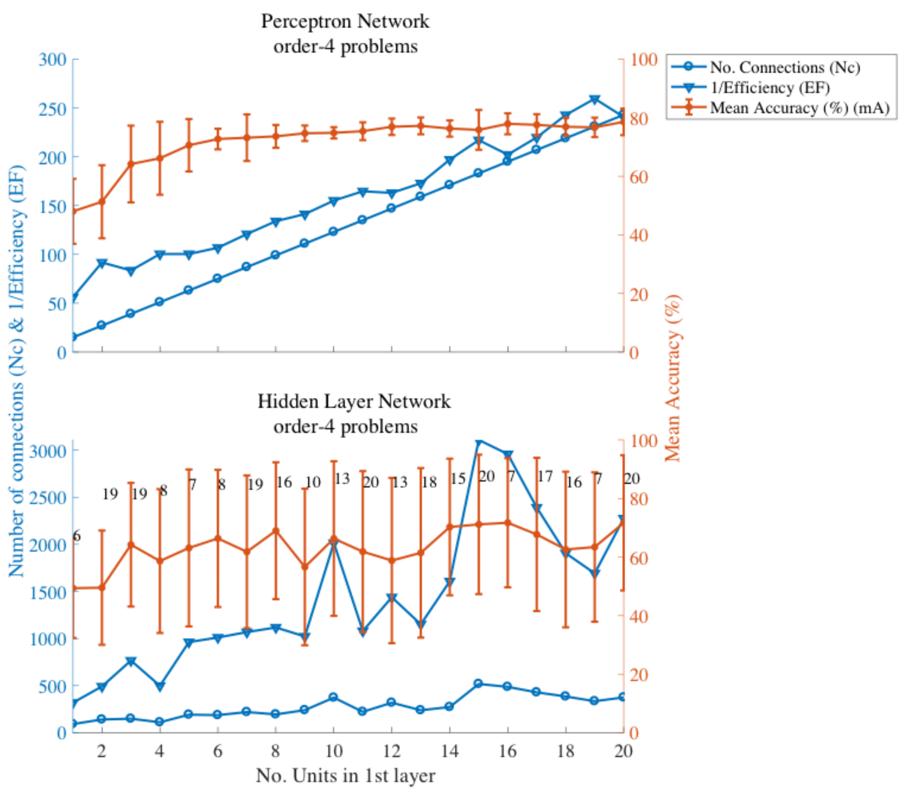

Figure 19 — Topology and performance of perceptron networks and hidden-layer networks in total accuracy (*At*), total number of connections (*Nc*) and Efficiency (*EF*) for order-4 problems

Table 19 — Preferential topology of network for order-4 problems

Network type	Numbered selected networks	Diff EF	Units (1st layer)	Units (hidden layer)	Connections
Perceptron network	1	-18,72	20	—	243
	2	-14,89	16	—	195
Hidden-layer network	3	-934,85	11	8	222
	4	-566,99	17	13	429
	5	-474.81	18	10	385

Table 20 — Performance of best network for order-4 problems

Network type	Numbered selected networks	Mean Accuracy (%)	Max Accuracy (%)	Epochs	Best Epoch
Perceptron network	1	63,82 (∓8,14)	75,68	173	167
	2	80,70 (∓3,45)	86,17	196	190
Hidden-layer network	3	61,84 (∓27,51)	90,91	322	311
	4	67,72 (∓26,16)	92,93	204	194
	5	62,58 (∓26,58)	90,91	202	192

For both orders of complexity problems (order-3 and order-4), perceptron networks seem to be more, or equally, reliable in terms of learning accuracy, more stable in terms of performance across trials, and to present a more linear evolution as units are added. Yet, perceptron networks for order-4 problems seem to not benefit as much from the addition of units. Hence, it is deduced that perceptron networks are more suitable for learning order-3 complexity problems, whereas hidden-networks are more suitable for learning order-4 complexity problems.

Finally, it is important to mention that the entire experiment was run several times with random selection of training cases and random initialization weights. Results remained similar across all simulations, including the same optimal number of units in the hidden layer per number of units in the internal layer (and the same levels of accuracy). Moreover, the same experiment was also run with the same number of training and test cases per order of problem complexity (N = 522), as well as the same number of falling left, balancing, and falling right configurations per order. This secondary test aimed at clarifying whether the different number of training cases had influenced the results, which it did not.

Discussion of Experiment 2

The connectionist structure required for concrete and abstract order problems is the same as found in Experiment 1. Differently, order-3 problems in a two-arm beam were solved with 100% accuracy, but not in a one-arm beam. This fact raises an interesting issue.

> For the formal order of complexity, where the operation is more complex, the increase in the number of problem dimensions favored learning.

Furthermore, while in a two-arm beam the optimal structure was a perceptron network, in a one-arm beam the optimal structure for solving formal problems was found to be a hidden-layer network. Even though, maximum accuracy of one-arm hidden-layer networks was approximately 85%. It seems that connectionist structures take advantage of having more input units to distribute the information until an output is generated, or, given less input dimensions, they require distributed connections across orders. This calls attention to an important aspect — there are properties of artificial learning models that might interfere with the properties of problems one wants to simulate, which stresses the importance of re-questioning how to better integrate these properties with artificial learning system's properties themselves. Namely, in humans, an increased number of problem dimensions is associated with an increased cognitive load, which decreases performance.

In regards to systematic problem-solving, it was found that both a perceptron network and a hidden-layer network learned to solve the problem with similar values of accuracy.

Accuracy was always below 86%, which shows that something else needs to change in the structure of the system such that high levels of accuracy are maintained.

The most relevant fact is that, in systematic problems, the efficiency curve shows that the addition of units in a perceptron network does not benefit learning, whereas the addition of units in hidden layer networks does. Second, even if a perceptron network was chosen as the optimal structure, the number of internal units would need to double as compared to the structure for formal problems.

In terms of transitions, although there was an interaction between the number of problem dimensions and the operations to be conducted, results did not change the nature of transitions across orders. In both experiments 1 and 2, the transition from concrete to abstract was memory-based (same number of units), and from abstract to formal was operationally-based (same number of layers, different number of units). Results also indicate that the transition from formal to systematic problem-solving is an operationally-based transition, according to the hypothesis. The idea that there are complexity jumps is kept, as has been repeatedly observed in behavioral experimental paradigms. Yet, the way the system needs to grow to account for more complex problems appears not to be linear. Even within operationally-based transitions, the system does not grow in a linear manner.

EXPERIMENT 3

The present experiment aims to study the interaction between the number of units and the connectivity pattern in neural network models as more hierarchically complex problems are solved.

Formal order-3 problems (stage 11) and systematic, order-4 problems (stage 12) were trained, separately. Lower orders 1 and 2 (stage 9 and 10) will not be included in this experiment because previous ones have shown that only one unit is required in the internal layer for the network to learn these problems with maximum accuracy of 100%. Hence, for order-3 and order-4 problems, five types of connectivity pattern networks were tested in hidden layer networks, two of which apply to perceptron networks, too (Figure 20, Figure 21,

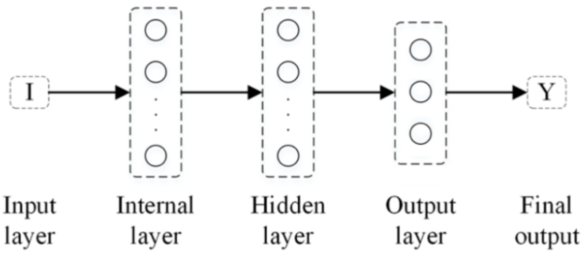

Figure 22, Figure 23, Figure 24, Figure 25, and Figure 26). It is hypothesized that a more densely connected architecture will require fewer units. Another alternative is that, with the same number of units and a denser connectivity pattern, the network will perform with higher accuracy. This hypothesis is based on the fact that the number of connections linking units grows throughout development (Geary, 2011; Johnson, 2011; Leite et al., 2016; Peters and Smedt, 2018), accounting for greater plasticity and learning abilities, due to a dynamic adjustment "within some fixed anatomical parameters" (Qin et al., 2014). The number of cases per OHC problem was equalized (N=522), where the percentage of output classes (falling left, balancing, falling right) was balanced.

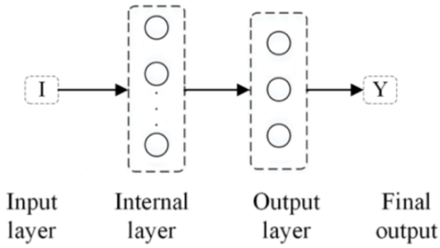

Figure 20 — Network 1p: Perceptron Network with Feedforward Connections

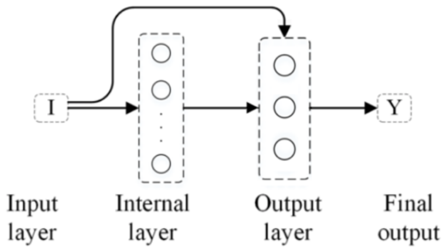

Figure 21 — Network 2p: Perceptron network with input connecitivity

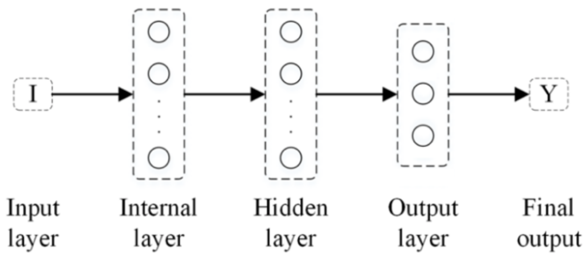

Figure 22 — Network 1: Hidden Layer network with Feedforward Neural Network

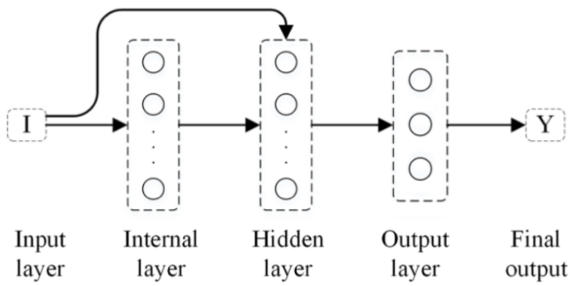

Figure 23— Network 2: Hidden Layer network with Input layer connected to the Internal *and* to the Hidden layers

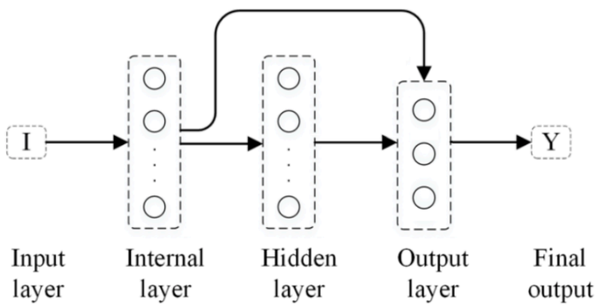

Figure 24— Network 3: Hidden Layer network with Internal layer connected to the Hidden *and* to the Output layers

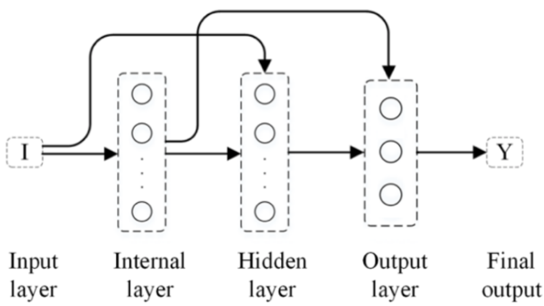

Figure 25— Network 4: Hidden Layer network with Input layer connected to the Internal and to the Hidden layers; and Internal layer connected to the Hidden and Output layers

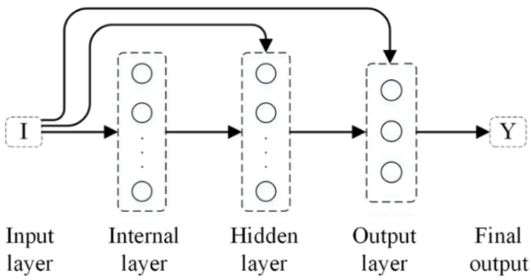

Figure 26 — Network 5: Input layer connected to the Internal, Hidden, and Output layer

Results of Experiment 3

In regards to problem-solving for order-3 problems (formal), perceptron networks with different types of connectivity patterns learned the problem with 100% accuracy. The feedforward connected network required 11 units, whereas the perceptron network with inputs connected to both internal and output layers required 9 units. Curiously, both networks required exactly the same number of connections, 135, to perform with 100% accuracy on the test set Table 21. If hidden-layer networks were used, three types of connectivity pattern networks performed with 100% (Table 22).

Table 21 — Selected perceptron networks for solving order-3 problems

	Units in internal layer	Units in hidden layer	Connections
Net 1p	11	—	135
Net 2p	9	—	135

Table 22 — Selected hidden-layer networks for solving order-3 problems

	Units in internal layer	Units in hidden layer	Connections
Net 2	1	7	103
Net 4	1	10	145
Net 5	1	4	88

In sum, formal problems were solved with a maximum accuracy by many types of networks, namely perceptron networks with different connectivity patterns and hidden layer networks with connectivity patterns sharing a common feature: all had input units linked to more than the internal layer of computations (here called input connectivity). The network with densest input connectivity (network 5) required the least number of connections.

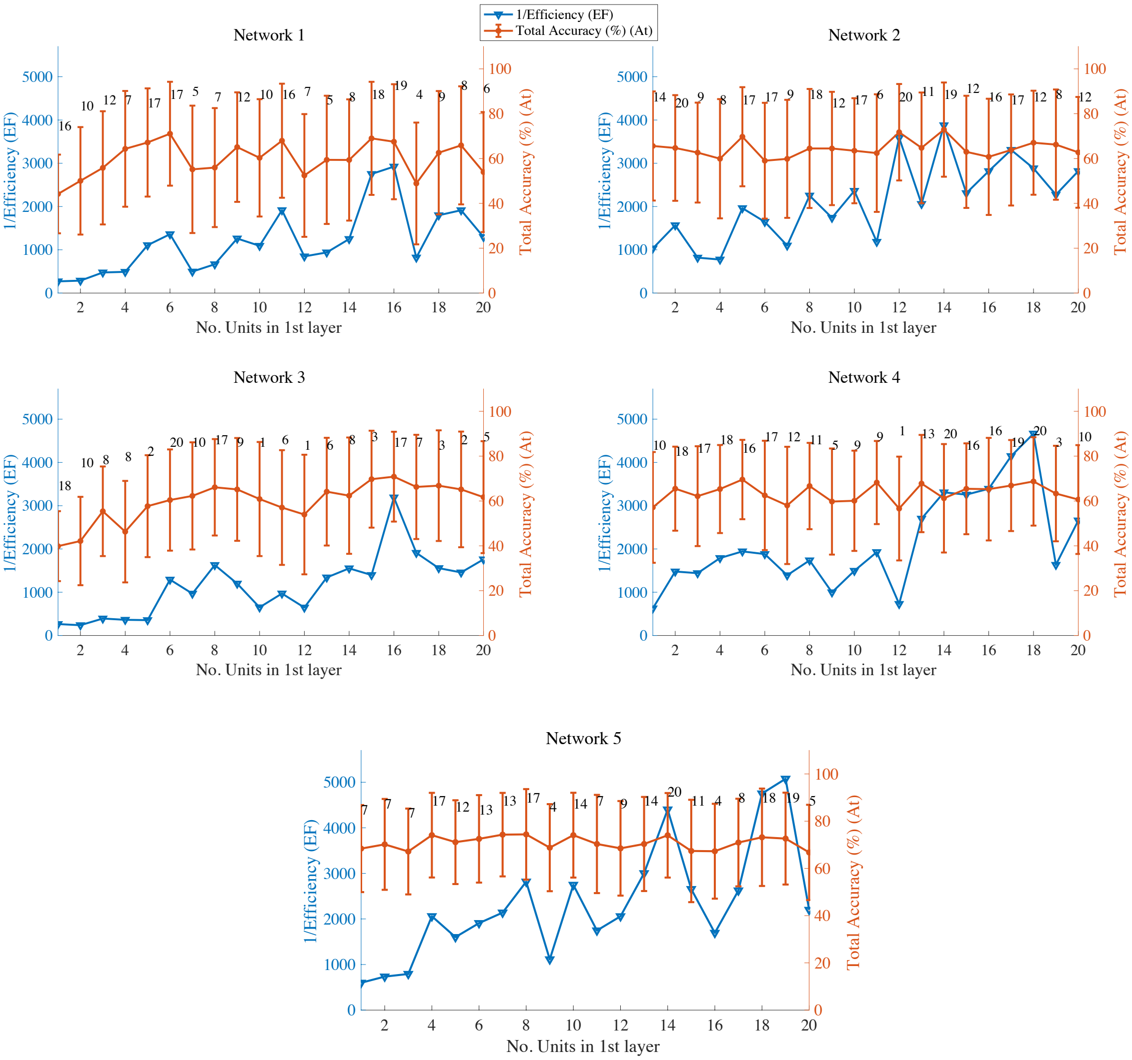

Figure 27 — Graphs representing the Performance and Efficiency of networks for solving Systematic problems

In regards to problem-solving for order-4 problems (systematic), all connectivity patterns showed a great variability in accuracy and efficiency (Figure 27). Nonetheless, as units were added in the internal layer, all network types showed an initial decrease in efficiency, followed by a first peak, local minimum. The network correspondent to this local minimum in efficiency has been selected. Table 23 contains the values that describe one selected network per type of connectivity pattern. Among all connectivity patterns, network 4 shows the best performance. Interestingly, again, two networks with different connectivity patterns (network 2 and network 5) converged to the optimal solution with almost the same number of connections (194 and 195, respectively), and both with inputs linked to forthcoming layers.

Table 23 — Structure of selected network per pattern of connectivity for order-4 problem-solving

Selected network	**Diff EF** ***(p – (p-1))***	**Diff EF** ***((p+1) – p)***	**Connections**	**Maximum Accuracy**	**Mean Accuracy (∓ sd)**
Net 1 (7-5)	-864.03	169.65	121	94.87%	55.18% (∓28.41)
Net 2 (1-14)	0*	536.11	194	92.31%	65.61% (∓24.34)
Net 3 (10-1)	-548.68	319.67	137	91.03%	60.87% (∓25.44)
Net 4 (12-1)	-1198.58	1972.13	171	96.15%	56.63% (∓23.15)
Net 5 (9-4)	-1705.64	1635.66	195	91.03%	68.79% (∓18.43)

* a value of zero indicates that there is no previous value of Efficiency, as this is the first network of the sequence.

Given the decrease in accuracy for solving systematic problems even with the addition of hidden units, we performed another cycle of network runs. We tested the same input connectivity, but added recurrent connections from the hidden to the internal layer. Yet, the addition of recurrent connections did not have an impact on the results.

Finally, given that all these networks are more densely connected than feedforward neural networks, it is important to mention that the number of connections in the selected networks does not surpass 1/3 of the number of training cases. So, results are likely not due to overfitting.

Discussion of Experiment 3

Results were clear in demonstrating the interaction between the addition of units and the presence of a determined connectivity pattern.

> They showed that when a determined connectivity pattern is present, performance increases. This connectivity pattern consists of linking input units to internal *and* to the forthcoming layers of computation: hidden and output layer.

These networks are here called networks with input connectivity. When this pattern was present, the best performance was found in networks that converged to the same total number of connections. Moreover, networks with input connectivity, either perceptron or hidden-layer networks, surpassed performance of previous simulations.

Formal problems, those where multiplication needs to be encoded to account for small torque differences, were solved with 100% accuracy. In what refers to perceptron networks, interestingly, networks with and without input connectivity recruited a number of units that resulted in the same number of connections. For hidden-layer networks, only those with input connectivity did learn the problem with 100% accuracy, but there was no tendency of converging to the same number of connections. The most important aspect was that the densest connectivity pattern recruited the least total number of units (8) and the least number of connections (88). For systematic problems, two of the three networks with input connectivity converged, again, to the same number of connections, approximately (194 and 195). The third one (with input connectivity) did achieve higher maximum accuracy above all. What seems to be most important is that, for the same connectivity pattern, the structure of networks for adjacent complexity problems seems to share commonalities: the number of units in the internal layer or in the hidden layer is the same, which indicates possible structural progressions. Below are three most relevant options for how networks structure might progress along increasingly complex stacks (Figure 28).

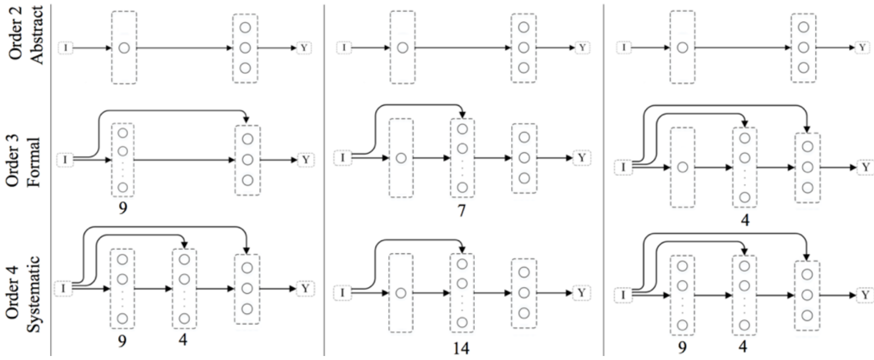

Figure 28 — Three possible progressions across stacks (option , 2 and 3, from left to right). Concrete-order stack is not represented as it contains the same number of layers and units as the abstract-order stack.

This suggests that different departing structures can be the starting point for higher-order structures.

Finally, results point out that dependencies among subparts (connections) have some degree of independency: independently of the number of units, optimal solutions of models converged to the same number of connections, which indicates that connections can be seen as components of the system, as well.

EXPERIMENT 4

Once different departing structures might self-organize to form higher-order structures (Table 24), the present Experiment 4 evaluated whether a higher OHC structure can be built out of the lower OHC structure, based on pattern-wise learning.

Pattern-wise learning is similar to layer-wise learning, but uses the weights of previous networks instead of the weights of previous layers. In other words, the present work tests whether all these possibilities of networks structures hold if pattern-wise learning is applied to the formation of higher-order stacks. The goal is to aggregate the results obtained so far and build a stacked generative architecture that develops from the concrete to the systematic stack, based on pattern-wise hierarchical learning.

Table 24 — Structural progression from concrete to systematic problem-solving

Option		**Order-1 network**	**Order-2 network**	**Order-3 network**	**Order-4 network**
1	*Network type*	Perceptron	Perceptron	Perceptron (input connectivity)	Network 5
	# Units	1	1	9	9 — 4
	# Connections	12	16	135	195
2	*Network type*	Perceptron	Perceptron	Network 2	Network 2
	# Units	1	1	1 — 7	1 — 14
	# Connections	12	16	103	194
3	*Network type*	Perceptron	Perceptron	Network 5	Network 5
	# Units	1	1	1 — 4	9 — 4
	# Connections	12	16	88	195

Also, pattern-wise learning is associated with the notion of hierarchical stability. According to a complex system's perspective, as a system develops, earlier lower-order structures gain higher rigidity and stability, whereas higher-order structures (emergent from the previous ones) are more dynamic and flexible. This means that the components of lower-order structures, as the system develops, change less than the components emerged at higher-order structures. This allows that the individual not only moves up to more complex stages, but also that lower-stages are protected from interference of higher stages. Hence, the individual also moves down to more elementary levels of performance, which confers on behavior great flexibility and ability to adapt to environmental circumstances (Fischer, 2008).

The weights of the components shared by the lower-order and the immediately higher-order stacks were kept as the initial weights of the higher-order stack. Yet, during learning of the higher-order stack, all weights were updated. According to a complex system's perspective, the allowance of weight update is in accordance with the fine-tuning of previous structures in order to accommodate higher-order information into more adaptive patterns.

Also, in order to approximate the idea that lower-order components become increasingly rigid and stable as the system develops, further tests were conducted.

In a subsequent step, different learning rates (LR) were applied. At each transition, the learning rate applied over lower-order weights decreased by 20%, 30%, and 50%. A final test was conducted in which the weight matrices composed by the same number of elements as in the lower-stack network was not allowed update.

Results of Experiment 4

Independently of the learning rate condition, the networks for order-2 problems always learned with 100% accuracy. Table 25 shows the performance of each three options of networks for formal and systematic problems. The LR conditions below 100% were aggregated in one single column of results because results were similar across these conditions. Each cell of the table contains the maximum performance accuracy of each network. The LR modifications did not impact the results in a significant manner. Yet, there is a tendency to decrease in accuracy as the LR is decreased, as expected.

Table 25 — Network's performance per structure and LR condition

Option	**LR = 100**		**LR < 100**		**No learn**	
	Formal Network	Systematic Network	Formal Network	Systematic Network	Formal Network	Systematic Network
1	98.72%	97.44%	97.44%	97.44%	97.44%	85.90%
2	94.87%	87.18%	94.87%	87.18%	93.59%	91.03%
3	96.15%	96.15%	96.15%	96.15%	98.72%	85.90%

Among all the three options, options 1 and 3 are the most stable in accuracy across OHC and the least impacted by LR modifications, as long as LR is different from zero. Option 2 suffers the greatest decrease in accuracy as pattern-wise learning is applied, as well as the greatest variability across orders of hierarchical complexity. Interestingly, options 1 and 3 share the input connectivity pattern, having inputs connected to all forthcoming layers of computation, independently of being a perceptron network or a hidden layer network.

Furthermore, the growth in the number of connections as the OHC of problems increased was plotted (Figure 29). The graph shows that option 3 approximates the shape of an exponential increase in the number of connections, as predicted by the MHC (Harringan and Commons, 2014). Curiously, option 3 corresponds to the network structural growth with the densest connectivity pattern, among those with input connectivity.

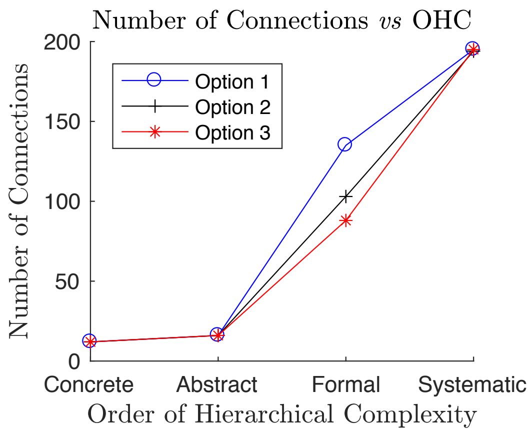

Figure 29 — Number of Connections vs OHC for each three options of network growth

Discussion of Experiment 4

First, structural growth is resistant to modifications in learning procedures. Second, the denser the input connectivity, the more robust the network is to learning procedures. The most suitable connectivity pattern was the one with the densest input connectivity, for two reasons. First, this pattern showed that networks remained stable in performance as the order of hierarchical complexity (OHC) of problems increased, practically independently of the modifications in learning rate (LR). Second, the growth in connections across stacks approximated the shape of an exponential increase (Harringan and Commons, 2014). These results are in line with neuroscientific results, which suggest that the brain's information processing capability gains more from increasing connectivity of the processing units than increasing processing units themselves and that, during development, there is an exponential increase in spike rates of synaptic activity (Brewer et al., 2009).

> Interestingly, this pattern of input connectivity has already been discussed as a very important finding for triangulating behavioral development, developmental cognitive neuroscience, and developmental cognitive computation.

The fact that pattern-wise learning was successfully applied and corroborated structural growth implies important considerations. It means that each stack is a good representative of the OHC, independently of the remaining stacks.

> It also means that the term “structural integration”, instead of “hierarchical integration” is well applied to the formation of increasing stages of development, or developmental stacks.

3. Global Discussion of Present Simulations

Hierarchy within a system cannot be traced by what is observable from the outside, but exists at the level of the structure. This was also true for connectionist representations of developmental problem-solving. Results indicate that the OHC is a strong criterion for characterizing stages and that connectionist models are strong candidates for capturing the developmental progression across stages, when stages are represented by OHC.

> The minimal structure of each connectionist model seems to be efficient and informative for representing stages of development.

Results also indicate that stack transitions might not be all of the same type, neither they make the system grow in a linear manner. Accordingly, opposite to the linearity that characterizes the OHC underlying stages of performance (from 1 to 17), neuroscientific methods have provided evidence of non-linear changes in structural architecture and functional organization in the developing brain. Hence, the OHC is represented non-linearly in the brain: what matters is the components of the system, how they are connected with each other (Johnson, 2011), and how they evolve as interactions with the environment proceed. This non-linear progression of connectionist structures was, as well, found in the present set of experiments, where the number of units and connectivity pattern among units varied.

Experiments 1 and 2 allowed for a distinction between 1) transitions based on an increase of the number of problem dimensions and 2) transitions based on a change of the operation required to coordinate problem dimensions. In other words, a memory-based transition occurred when the number of problem dimensions increased, but the operation to coordinate dimensions remained; and an operationally-based transition occurred when the number of dimensions remained, but the operation to coordinate those dimensions changed. This was true for the two types of beams, a one-arm beam and a two-arm beam. It was found that an operationally-based transformation underlies the passage from abstract reasoning to formal reasoning, representative of higher-order cognition, unique in humans (Commons et al., 2014c; Leite, 2016). These difference in the nature of transitions matched the different transitions in this particular simulation scenario, from concrete to abstract and from abstract to formal, and showed, again, that connectionist models capture well cognitive phenomena.

Yet, there is an important question attached to this interpretation. On the one hand, the summing operation (abstract stage 10) can be understood as a horizontal extension of the operation of counting (concrete stage 9). Specifically, in the representation of counting and summing, what changes is not the operator "+", but the parcels that are to be associated by the operator. Differently, in the transition from abstract (stage 10) to formal (stage 11), the operator changes from "+" to "×", independently of the parcels. The fact that the operator does not change from concrete to abstract might underline the fact that there is not an operationally-based transition here. On the other hand, the first computation performed inside the units of a neural network is a weighted *sum*. This impedes that the operation of summing is approximated by a composition of units, as it is already explicit rather than approximated. Hence, this property of connectionist models might impede to consider the passage from concrete to abstract as an operationally-based transition. In other words, the fact that artificial neurons in a neural network already sum the incoming inputs might give away a lot of the operational structural transition between counting and summing. This might be leading to incurring a type II interpretation error. An operationally-based transition might exist in some way from counting to summing, which artificial neurons did not capture. This limitation is indirectly associated with limitations of previous simulations, where the assimilation function was set to product, so as to adapt to multiplicative problem-solving (Reyes et al., 1997).

Nonetheless, results continued showing that, once the system is required to transit to formal *and* postformal stages, the nature of transitions goes from memory-based to operationally-based, at least in the present simulation scenario. The emergence of the cognitive potential to *operate* with abstract information (formal 11 and postformal stages 12) instead of only creating abstractions (abstract stage 10) eventually points out a major factor underlying developmental jumps, even evolutionary jumps. In fact, the formal stage of reasoning is assumed to be evolutionarily distinct, responsible for higher-order cognition, and only present in humans (Leite, 2016).

> Findings suggest that a formal reasoning implies a shift from memory-based strategy for problem-solving to an operationally-based strategy, in accordance with higher-order cognition.

Based on these results, there might be three major classes of development, or units of analysis. The first occurs during infancy, where the baby learns to recognize and coordinate sensations, perceptions, and actions, until representations are formed that allow the child to form concepts. These are called perceptual groupings or representations (Sloutsky, 2010). This class has not been object of the present work. The second class takes place afterwards, when the child becomes able to inter-relate representations, forming concepts, and inter-relating concepts with other concepts, as they appear to be inter-related throughout the course of *experience*. This second class lasts until the child is able to create abstractions, or variables (abstract stage 10), which are the result of a sequence of similar experiences (Sloutsky, 2010). During this second developmental phase, memory-based transitions might be dominant. Once the child begins to use variables, a new unit of analysis has been "reached". At the third class, the child, now an adolescent, starts to use variables as objects and relate them together, making use of formal (and postformal) reasoning. At this third unit of analysis, the individual focuses on unobservable data features or unobservable inter-relations between them, reason why this latter acquisition largely depends on maturational abilities of higher-order cognition (Sloutsky, 2010). This third phase might be strongly associated with operationally-based transitions. This does not mean that each type of transition is exclusive of each phase.

Moreover, in Experiments 1 and 2, the dropping in accuracy values of networks that solved increasing OHC problems let one guessing that something else should be tested. Experiment 3 shows that there is a strong and beneficial impact of the connectivity pattern of the networks on their performance, specifically in the case of input connectivity. Experiment 3 also showed that if problems are segregated by OHC and if there is a network per OHC, each network performs better than a network built to solve all complexity problems of a set. Here, for the first time, a network with summing as the assimilation function was able to represent the multiplication operation and solve formal problems with 100% accuracy.

The influence of connectivity pattern, besides the impact of the number of units, is relevant for a comparison with the natural learning system. In fact, one the one hand, a fixed number of neurons is correlated with the mean stage that a species achieves, being it the stage of formal operations in humans (Commons et al., 2014b). On the other hand, the number of connections changes throughout development, which points towards a dynamic adjustment of the neural architecture within some fixed anatomical parameters, as experience proceeds (Qin et al., 2014). In line with this, the densest input connectivity pattern has been associated with the strongest resistance to learning modifications across stacks, to better performance, and to greater biological plausibility due to the number of required units throughout stacks.

The influence of connectivity pattern is further relevant for a comparison with the cognitive arrangement (Leite and Rodrigues, 2018). Namely, Chapter II, Section C presented a preliminary experiment that defended that a higher-order output is the result of reorganizing the inputs as more complex operations are conducted. It showed that it is not the lower-outputs that become object of higher-order outputs, as has long been advocated by many stage theories. This implies that inputs, or percepts, must always be used to generate increasingly complex outputs (Leite et al., 2018). This is in line with cognitive and neuropsychology studies, too. These have long revealed that the attentional system is the cognitive sub-component precursor of higher-order cognition, in that it selects the environmental features to be operated upon (Sweller, 2004). Hence, the impact of the input connectivity pattern further allows for suggesting that connectionist models might even be able to capture the horizontal organization of the cognitive apparatus, leading way to coordinate the horizontal and hierarchical nature of information processing. It has not been possible yet to determine the rule for stack transition, but results point towards the necessity of growing input connections at increasingly complex structures.

Still from the results of Experiment 3, it was possible to identify that a higher-order structure included the components of a lower-order structure. The fact that different departing structures can be the starting point for higher-order structures highlights that connectionist models eventually capture individual variability *within* structural progression. Interestingly, this was only true with networks with the same connectivity pattern. This is worthy to mention because networks per OHC problem were trained separately and the degrees of freedom brought about by the choice of hyper-parameters could have made this progression very difficult, even impossible, to detect. This means that connectionist models, if treated as information processing models, fail to represent and solve problems in a developmental way. Differently, if connectionist models are allowed some room to represent problems their way (as long as problems are segregated by OHC), they actually capture the principles of complex systems and approximate the laws of cognitive and brain development. Results, thus, allowed for proposing a different method for developing a self-organized cognitive architecture, based on evolving stacks.

> The dominance of the connectivity pattern in the structural progression across stacks opens up the way for building further possible hypotheses that relate connectionist models and natural systems.

The notion of endo-causality (Maturana and Varela, 1928) implies that a system behaves based on some form of internal causality, where its sub-parts form dependencies with each other. Also based on the notion of endo-causality, the present study suggests that dependencies among subparts (connections) have some degree of independency, as independently of the number of units, optimal solutions of models converged to the same number of connections in formal problem-solving. Hence, although the mechanism of HI, as is defined in the MHC, seems not to be the way of creating a hierarchy of problem-solving, results still point towards the idea of a self-fed structural hierarchy. One might talk about "hierarchical structural integration", where "the higher-order structure is formed out of the lower-order structure".

In sum, the method here elaborated for studying and implementing development in a connectionist model was corroborated by the results of these four experiments. Connectionist models captured stages of development, stage transitions, and developmental progressions. They did this by approaching biological plausibility, namely by being influenced by the same parameters that underlie stage representation in the brain — number of neurons and pattern of connections. They did this also by showing the possibility of modelling individual variability, given that different structural progressions are possible. Hence, it is concluded that the proposed method is a good starting point to study how developmental transitions can be ascribed and represented in connectionist models, not to mention that its applicability to the neuroscience domain can bring about valuable information.

For now, in order to build a stacked developing connectionist structure that approximates and represents development, two conditions need to be met.

First, each stack needs to be triggered by a given set of problems a priori characterized. This is important such that a lower-order structure never gets to be substituted by a higher-order structure. If the characteristics of the problem re-activate a given structure, one can be certain that lower-order structures are always available to the global system. Second, each higher-order stack can be built out of the lower-order stack. This is important such that the global model can be said to "mature", or to develop on the basis of a self-organizing force. If both conditions are true, then, higher-order stacks will always be generated by lower-order ones, and the global model will always be able to re-activate lower-order stacks, which means protecting simpler representations from the interference of more complex ones (hierarchical stability).

CHAPTER V

CONCLUDING REMARKS

The present Chapter V is based on the joint considerations and results obtained throughout this dissertation. It exposes the concluding remarks of the work conducted for the topic under investigation — simulating cognitive development in a connectionist model — including relevant findings, limitations, and advantages. This chapter also summarizes the main premises for the "Foundation of a Hierarchical Stacked Neural Network model for Simulating Cognitive Development", highlighting that the major contribution here provided is a pot of ideas and methodological procedures that hopefully allow for a new tripartite research path to emerge.

CONCLUDING REMARKS

Ascribing developmental properties to an artificial system has been here associated with approaching an increased hierarchical flexibility and stability, which is a problem never fully answered before. Above all, this dissertation provided a method for doing so — ascribing developmental properties to an artificial system — and conduced a set of experiments that corroborated the method itself and the possibility of building a developing connectionist model. Furthermore, as mentioned in the beginning of this dissertation, the main goal of Artificial Intelligence (AI) can be used both 1) as a method for testing models of information processing and 2) as the implementation of what is (provisionally) known about information processing in a living organism. Here, the main premises of the Model of Hierarchical Complexity (MHC) have been tested in order to implement an artificial system that learns in a developmental way. The methodological considerations and obtained results can be viewed bi-directionally. They show that some theoretical premises of the MHC should be revised, and that some should be kept in order to set the foundation of a developing system — a Hierarchical Stacked Neural Network model that simulates cognitive development. The following set of shortly answered questions and issues aims at summarizing the findings of the present dissertation, as well as their backwards implications.

1. Bridging concepts across fields

Theories of behavioral development have long been used to describe, eventually explain, how individuals produce more adaptive functional and behavioral patterns as they mature in cognitive capacity. The transduction of this adaptive movement to a machine has been analogized and operationalized as flexibility and stability. Flexibility concerns the machine's capacity to solve diverse problems, whereas stability concerns the machine's capacity to form and fix stable representations of similar problems. Yet, the criteria that discriminates the end of a certain type of problem and the beginning of another are not totally clear, which prevents the clear operationalization and implementation of flexibility and stability. In turn, the result is that machines do not go beyond strict programmed constraints of learning and can only learn the narrow set of tasks they are programmed for and presented with.

The present dissertation began by highlighting the relevance of simulating learning *and* development in an artificial system that aims to reproduce human cognition. From a psychological and cognitive point of view, learning and development are two intertwined, but discriminated products of interacting with the environment. Whereas learning is a continuous and substitutive process through which error rates decrease at a certain task performance, development is a discontinuous process, where successive and cumulative patterns of problem-solving get more complex. Hence, during learning, the functional and behavioral patterns are updated and substituted by more adaptive ones. During development, the functional and behavioral patterns that initially only solved simpler problems are reorganized, not substituted, and form more complex patterns.

By discriminating learning and development, the operationalization of flexibility and stability were discriminated as horizontal or hierarchical. Horizontal flexibility and stability concern the formation of many representations of the same complexity. Hierarchical flexibility and stability concern the formation of increasingly complex representations. Both horizontal and hierarchical flexibility and stability show a similar progression across individuals, which points towards the existence of a structural criterion for discriminating problems. The present dissertation focused on the hierarchical flexibility and stability of a connectionist learning model. In other words, it focused on ascribing to it developmental, or maturational, properties.

2. What are we talking about when we talk about the "maturation" of an artificial system?

In developmental psychology terms, maturation refers to the developmental properties of individuals, who progress along a developmental stage sequence. In developmental cognitive neuroscience terms, maturation refers to changes that occur at a neuroanatomical and physiological level, which correlate with behavioral developmental spurts. Hence, ascribing maturational properties to an artificial learning system is a synonym of representing the biological factor underlying stages of development. The structure of the brain is defined in terms of its components and connections; the structure of an artificial neural network is determined by its components and connections; if the structure in the human brain changes to accommodate development; the structure of an artificial neural network must change to accommodate development, too. These changes *are* the maturation of the system.

Developmental behavioral psychology is somewhat divided into domain-general and domain-specific theorists. Domain-general developmental theories provide evidence that individuals progress along a one-way sequence of development that is similar across domains of knowledge. They strive for identifying the universal factor that describes and explains cognitive capacity across domains. Differently, domain-specific theories assume that developmental sequences differ across domains of knowledge. However, because domains of knowledge overlap with each other, developments in one domain foster developments in other(s) domain(s), explaining the joint developmental growth across domains. Developmental neuroscience has focused mainly on domain-specific development as it is easier to a priori select and identify the particular regions of interest, given the particularities of the task (language, arithmetic, memory-load, etc.). However, building a system that matures implies that it is imbued with a representation of a universal cross-domain developmental factor, which is structural, invariant and systematic. If this universal factor is evidenced, then, it certainly ascribes a necessary structure to the way individuals interact with such a changing and complex environment. For that, a domain-general developmental theory was chosen to delineate the theoretical premises underlying this work — the Model of Hierarchical Complexity (MHC).

3. The Model of Hierarchical Complexity: Horizontal versus Hierarchical Complexity

The Model of Hierarchical Complexity (MHC) was chosen as the domain-general theoretical reference for a number of reasons. The is a domain-general theory that postulates that individuals go through successive stages of development as they mature. At each stage, they become able to solve increasingly complex problems. One of the contributions of this theory is that it formalizes the existence of two axes of complexity for characterizing problem-solving abilities: a horizontal axis and a hierarchical axis.

The horizontal axis represents the learning process that, during maturation, capacitates the individual to solve tasks with an increased cognitive load. An increasing horizontal complexity corresponds to an increased memory load of a task, yet, it is not what characterizes the transition for a higher stage of development. This is also associated with the horizontal flexibility and stability of a system, which creates representations of several features of the environment. The vertical axis, on the contrary, represents the developmental process that capacitates the individual to solve hierarchically more complex problems. The vertical axis is associated with hierarchical flexibility and stability of a system, which uses the horizontal representations formed so far to create a more complex representation of the environment. The Order of Hierarchical Complexity (OHC) is the variable represented across the vertical axis. It is a unidimensional abstract measure of development that mathematically formalizes the existence of the universal, structural, and systematic factor underlying stages of development. It predicts with high accuracy the performance of individuals. The OHC was the criterion for setting the hierarchical stability of the system, given that it homogenizes performance within stage, across domains. The MHC also presents evidence that this universal factor is imprinted in the neural architecture, allowing for bridging developmental psychology and developmental cognitive neuroscience.

Another advantage of choosing the MHC is that it departs from evaluating the complexity of the task to be solved, or problem. The OHC is the property of problems that interacts with individual's cognitive abilities, such that some problems are solved before others. If the individual, or agent, correctly solves the task, it is assumed that the stage of development of the solver is the same as the OHC of the task. Hence, this allows for assessing behavior in a structural, non-mentalist manner. More importantly, it allows for evaluating behavior of humans and non-humans, including machines, and for comparing behavior across species and across agents, including human-machine comparisons. Finally, the MHC assessment method has been applied to a number of human developmental tests, such as the balance scale test, allowing for a comparison between its domain-general assessment method and other domain-specific behavioral assessment methodologies, relevant for the present case. The balance scale test is the scenario from developmental psychology borrowed to guide simulation experiments.

4. How are stages represented?

The representation of stages depends upon the OHC of tasks, or problems, to be solved. The OHC has been evidenced to be traced back in the neural architecture at birth. Hence, a study on the neural correlates of development is important to identify the possible bridging between developmental cognitive neuroscience, and developmental connectionism, given that there are structural similarities between a neural network model and the brain. A methodological proposal has been delineated to identify how a networked system represents the OHC, with validity for both fields. The difference is that, while the brain has an a priori representation of stages within its patterns of activation, in connectionist models one has to find the most suitable pattern for representing each stage and transitions across stages. In terms of neural network stacks, the number of units and the connectivity pattern among them have been here experimented.

Furthermore, the MHC assesses very clearly the OHC of individuals, given their performance on a number of developmental tests, but it poorly describes what is at stake in the determination of the OHC. Although the OHC is assumed to be a unidimensional and abstract measure, is has been associated with three possible factors — the number of problem dimensions, the type and number of coordinated operations, and the type and number of generated actions, or outputs.

5. How are stage transitions processed and represented?

It is a fundamental aspect to consider in the simulation of development in a connectionist model: how the system transits from one stage to the next The MHC also describes, along with other developmental theories, a mechanism for stage transition: the mechanism of hierarchical integration. It postulates that lower-order outputs are coordinated by means of a certain operation to form higher order inputs, which, in turn, form higher order outputs. However, the first experimental study of this dissertation did not validate this mechanism. It pointed out that stage transition lies at the operations that are conducted over the problem dimensions, but that the original problem dimensions need always to be taken into account to form higher-order outputs. Still, if OHC relies on the number of problem dimensions and on the operation coordinating them, complexity has a composite representation, which requires clarification.

In order to represent stages and stage transitions, a new method was proposed. A minimal structural complexity artificial model was created to solve each OHC problems, separately and independently. Then, the structure of models of adjacent OHC was compared. This allowed to address two important considerations of sage transition. First, that each stage is relatively independent from the previous and acts as an attractor of the system. Second, that each stage is supposedly formed out of the previous. In the present experiments, by not imposing structural constraints on higher order networks, the most likely relation between adjacent networks was the OHC. Also, by not imposing any a priori relation between the structures of models for each OHC, the generation of a higher stage based on the previous was tested anew. If there was a structural relation, then, it would likely derive from the representation of the OHC in connectionist models. Results showed that the structure of a higher order model can actually be built out of the structure of the lower order model, even departing from different structures. The term Structural Integration was proposed to substitute the term Hierarchical Integration. Furthermore, two types of structural transitions were found, which questioned whether stage are equally spaced in terms of hierarchical complexity.

6. Relevance of using connectionist models

Connectionist models, or artificial neural networks, are the artificial systems of reference o for a number of reasons. First, they have been created to approximate how the basic units of the central nervous system, the neurons, operate. Second, the structural complexity of connectionist models is based on how many units there are and how they are arranged by means of connections, where connections represent synapses. These models allow to establish the following parallelism — the connectivity of a neural network determines its structure; the connectivity in the human brain also determines its structure; the connectivity in the human brain changes to accommodate development; the connectivity in a neural network must change to accommodate development, too. Second, connectionist models have introduced a learning procedure based on a composite continuous function, which continuously substitutes the non-adaptive weights attributed to connections by more adaptive ones. This highly approximates the continuous and substitutive nature of learning observed in natural systems. Third, it has been proved that a sufficient number of units connected and distributed across layers of computations approximates any complex mapping between inputs and outputs, which makes these models potentially good learners for every task.

7. Difficulty in simulating development in connectionist models

However, connectionist models pose a number of problems for simulating development. The task of coordinating learning and development in a biologically plausible model, such as neural networks, is not straightforward. First, the discontinuity of developmental jumps threatens to corrupt the differentiability and continuity upon which connectionist learning algorithms are devised. Second, the formation of new developmental patterns, or connectionist structures, threatens the hierarchical stability of the system. This means that lower-order patterns are not protected from the representations created at higher-order patterns, due to continuity in learning. This problem has been called the problem of stability, catastrophic forgetting, or catastrophic interference. In the present dissertation, it has been called the problem of *hierarchical* stability, given it is a matter of stability concerned with the formation or more or less complex representation of a certain problem.

8. How to coordinate learning and development in a connectionist model?

The idea underlying this dissertation is the creation of a connectionist model built in stacks, or partially independent connectionist structures, in a way that 1) each higher order structure is developed based on the components of the previous and 2) whenever a lower-order complexity problem is presented to the model, it reactivates a lower-order structure, which is protected and has not been substituted during the developmental process. This idea carries on a third difficulty. Because connectionist models are universal mappers, there is the risk that each stack, or structure, is not the minimal complexity structure for solving that particular OHC problems. This is similar to saying that there is the risk that each stack does not only represent the factor that accounts for hierarchical stability. Hence, the objective is to find the minimal connectionist structure that accurately represents and differentiates OHC computationally.

9. How do the developmental properties of problems interact with the properties of artificial learning models themselves?

Results showed that connectionist models represent differently the impact of the increase in the number of problem dimensions and the increase in hierarchical complexity of the underlying operation. Two types of problem dimensions were tested — the dimensions that participate in problem-solving (weight and distance), and the dimensions that participate in simulating the problem (a one-arm and a two-arm beam). In both a one-arm and a two-arm beam, there were problems with differing number of dimensions (weight and distance). Namely, memory-based transitions occur when weight and distance become coordinated, but the operator to coordinate both sides of the beam remains, "+". In these transitions, the number of units composing the model does not change, but the number of used connections does. Operationally-based transitions occur whenever the operation needed to change. In these cases, the number of units and layers composing the model changes, as well as how units are best connected. Operationally-based transitions seem to characterize the passage to formal stages, characteristic of higher-order cognition.

When comparing the performance of a one-arm and a two-arm beam, transitions remained the same throughout OHC, but an increased number of problem dimensions (for a two-arm beam) improved networks performance. This is opposite to natural systems. As the number of problem dimensions increases, problem-solving faces an increased processing load, which decreases performance. This is an example of how the properties of the artificial system interact with the simulated properties of a natural system, in what concerns information processing. Another example is the possibility that the structural transition between counting (order-1) and summing (order-2) might be given away, since artificial neurons in a neural network already sum the incoming inputs. In order to disambiguate this effect, further simulations are necessary.

In sum, operations conducted over the inputs seem to dominate in the face of an increase in the number of inputs dimensions. These operations seem to be imprinted in the connectionist structure, implying changes in the number of units, layers, and in the connectivity pattern among units. Problem dimensions do not impact this structure, but they impact the ability of the model to solve problems more or less accurately.

10. How does a neural network model represent operations of increasing OHC?

Results have shown that each stack was able to represent the OHC independently of the other stacks, by varying the number of required units, layers, and connectivity pattern among units. A relevant tendency was found. Selected optimal network structures with different connectivity patterns, for the same OHC, tended to converge to the same number of connections (independently of the number of units). This is curious and important, ascribing some independency to the connections. They are not mere links between the components of the system, but can be viewed as components themselves.

Importantly, there was a connectivity pattern that seems to be more efficient for representing all OHC and more biologically plausible than the remaining—the input connectivity pattern. This means that the units in the input layer are connected to the units in all forthcoming layers (internal and hidden; internal and output; and internal, hidden and output layers). The third and densest input connectivity pattern showed the best accuracy, the strongest resistance to learning rate modifications, and an increased biological plausibility. Input connectivity underlines the fact that lower order inputs are reorganized by means of higher order operations to form higher order inputs, which in turn will form higher order outputs.

11. How does a neural network model represent the progression of OHC?

Results showed that adjacent OHC stacks capture a structural developmental progression, as long as structures share the same connectivity pattern. This structural progression led to suggest that the process of stage transition might be called "structural integration". This was possible to test, instead of imposing, due to the employed method, which segregated the problem space into disjoint subsets of OHC and then trained several independent networks for each subset to evaluate the best fitting structure. Not only a structural progression was found, but it was also found that each higher-order network could be trained based on the components of the lower-order network. The progression across OHC has been represented by a process called pattern-wise learning. Pattern-wise learning is similar to layer-wise learning, but the elements that remain fixed from one stack to the other are not necessarily disposed along the same layer. Pattern-wise learning means that the new structure, or stack, is built out of the elements of the previous stack. It is important to underline that structural integration applies to different departing structures, as long as the adjacent structures share the same connectivity pattern. Hence, these findings allowed for determining that connectionist models capture a progression in adjacent OHC and eventually can capture individual variability in the developmental process, but more network parameters need obviously to be tested and combined.

12. What considerations are worth mentioning for simulating development in a developing connectionist model?

Two of the main difficulties of ascribing maturational and developmental properties to a connectionist model are 1) the corruption of continuity and differentiability of the learning function, and 2) the protection of lower patterns of information processing (hierarchical stability). The present dissertation could answer both difficulties. The first was answered by the implementation of a generative architecture, that ascribed developmental and qualitative spurts of information processing procedures to an artificial neural network that were not a priori constrained. The second difficulty was addressed by the fact that the emergence of new network components was driven by a fixed criterion imposed over the input set — the OHC. With that in mind, new components that emerge for solving higher-order complexity problems are associated with that particular OHC and can be anytime retrieved or deactivated. This means that if a lower OHC problem is eventually presented to the network, the network is able to apply the minimal complexity structure to solve that problem. Lower order levels are protected from upwards interference. This is the property that confers on the natural system a great capacity to adapt — the ability to move up to more complex levels, as well as to move down to more elementary levels of information processing.

13. What methodological factors limit the scope of the present results?

A number of limitations are attached to the conducted studies. First, the hyper-parameters were heuristically chosen, which leaves plenty of room for variability in the results if other choices have been made. However, the results seem to be consistent throughout the entire set of studies, suggesting that the chosen independent factors are not strong enough to interfere in what has been observed. Second, the findings can only be said to apply to the balance scale test, although the OHC is a domain-general variable that explains and predicts behavior across domains of knowledge. Third, the search for the best network is ill-defined. A visual inspection of the results was sometimes necessary to select one or another network among all those that were trained. Hence, an improved and automated procedure for the selection of the best network is necessary. A closer inspection of the final weights of each OHC is necessary, such that in denser connectivity pattern networks, there is a process of pruning. This is informative to the process of structural progression, but has not been conducted in the present work.

14. What methodological factors benefit the scope of the present results?

This method also presents a number of advantages. First, it is a biologically plausible method in several forms: 1) the segregation of the problem space per subsets of problems of adjacent OHC has been confirmed to underlie how the human brain perceives and responds to information; 2) there is neuroscientific developmental cognitive data that supports the generation of increasingly complex structures as the brain matures and the individual grows in stage of development; and 3) even though the selection criteria is yet to be clearly defined, the fact that the best structure is not necessarily the most complex structure for a given OHC is in accordance with the pruning of synapses during development. There is an initial overabundance of synapses followed by a steep reduction: some synapses remain active and some are eliminated. Second, this method allows to verify how is it that a connectionist model represents stages of development rather than to merely reproduce what is observed in human behavior. Third, this method has applicability in two distinct, but related, fields of study — developmental cognitive neuroscience and artificial intelligence — which facilitates a bidirectional, interdisciplinary, interpretation of findings. It is important to test the same phenomena at different levels of observation because a wide variety of theories can accurately account for data collected at one level of observation, namely cognition or developmental psychology. If the same phenomena is tested across fields and methods, then, “there is a more robust account of that same data due to the overlap of constraints between the methods involved" (Johnson, 2011).

15. How can this dissertation inform about future work?

As mentioned in the beginning of this dissertation, it is beneficial (and necessary) that general theoretical and methodological concepts and principles are defined such that the work that is conducted at smaller scales at different specialized laboratories can be scaled up. In what concerns the field of AI, a shared definition of fundamental concepts and methods allows for raising the *qualia* of AI as a field running over its own rules instead of running over the rules of the fields which created it. This eventually renders AI to become a cross-paradigm, rather than a set of overlapping paradigms.

The major contribution of the present work is the possibility that a research line emerges that triangulates developmental psychology, developmental cognitive neuroscience, and computational cognition, specifically artificial neural network models. This triangulation is important such that a common research entity is construed in the overlap of so closely communicating disciplines, strengthening the scope and validity of results and interpretations. In practice, this dissertation proposed a method that has direct applicability in the fields of developmental cognitive neuroscience and artificial intelligence, particularly in connectionist models. The idea of this method was to approach an old problem that never saw a robust solution — the problem of coordinating flexibility and stability, along with the problem of coordinating learning and development.

In the particular case of building an artificial system that learns in a developmental way, there are many unanswered questions. For instance, 1) what other factor and network hyper-parameters influence the representation of OHC per stack? This seems to be a never-ending search. Furthermore, the representation of the OHC goes along with the ability of the system to identify the OHC of a given problem, allowing to activate lower order stacks, if needed. 2) Does the nature of stack transitions hold for lower order and higher order transitions? Actually, this simulation scenario only allowed for studying a narrow segment of the developmental sequence, which deserves further testing, from the lowest to the highest

OHC. 3) Is this method applicable to the simulation of other developmental scenarios? This is a most useful question, as one of the reasons why the MHC has been chosen is the fact that it is a general stage theory. The advantages of general stage theories for understanding and simulating development have been pointed out. Hence, the external validity of the MHC within the scope of AI needs to be tested. 4) How to ascribe autonomy to stack transition? Autonomy to the stack transition process is the ideal finding. It means that the previous questions have been answered. It also means that the system autonomously reproduces U-shaped learning curves as every transition occurs: as a new stack is initialized. Moreover, it means that another research question can emerge, concerning the biological and experiential factor that triggers the adaptive upwards movement, namely, the error, creativity, curiosity, etc. A last question here highlighted is 5) how to limit the hierarchical complexity that an artificial system achieves, based on the biological factor that limits stage? Is there a limit?

BIBLIOGRAPHY

BIBLIOGRAPHY

Abellán, F.J., Calvo-Llena, M.T., Rabadán, R., 2015. Harmonic Scale of Development. A proposal of integration by which to assess child development. An. Psicol. 31, 837–848. https://doi.org/10.6018/analesps.31.3.199901

Alnajjar, F.S., Yamashita, Y., Tani, J., 2012. Static and Dynamic Memory to Simulate Higher- Order Cognitive Tasks. Presented at the International Joint Conference on Neural Networks, IEEE, Brisbane, Australia. https://doi.org/10.1109/IJCNN.2012.6252749

Anand, R., Mehrotra, K., Mohan, C.K., Ranka, S., 1995. Efficient classification for multiclass problems using modular neural networks. IEEE Trans. Neural Netw. 6, 117–124. https://doi.org/10.1109/72.363444

Anderson, J., Carter, C., Fincham, J., Qin, Y., Ravizza, S., Rosenberng-Lee, M., 2008. Using fMRI to Test Models of Complex Cognition. Cogn. Sci. 32, 1323–1348. https://doi.org/10.1080/03640210802451588

Anderson, J., Fincham, J., 2014. Discovering the Sequential Structure of Thought. Cogn. Sci. 38, 322–352. https://doi.org/10.1111/cogs.12068

Armbruster-Genc¸ D.J.N., Ueltzho¨ffer, K., Fiebach, C.J., 2016. Brain Signal Variability Differentially Affects Cognitive Flexibility and Cognitive Stability. J. Neurosci. 36, 3978–3987. https://doi.org/10.1523/JNEUROSCI.2517-14.2016

Arnold, L., Rebecchi, S., Chevalier, S., Paugam-Moisy, H., 2011. An Introduction to Deep Learning., in: Proceedings of the European Symposium on Artificial Neural Networks (ESANN). Presented at the European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium.

Asaro, P.M., 2008. From Mechanisms of Adaptation to lntelligence Amplifiers: The Philosophy of W. Ross Ashby, in: The Mechanical Mind in History. MIT Press, Cambridge, MA, pp. 149–184.

Bair, D., 2004. Jung: A Biography, 1st ed. Little, Brown and Company, Boston.

Banks, M.S., Ginsburg, A.P., 1985. Infant visual preferences: a review and new theoretical treatment. Adv. Child Dev. Behav. 19, 207–246. https://doi.org/10.1016/S0065-2407(08)60392-4

Bartolotta, T., Shulman, B., 2013. Child Development, in: Language Development: Foundations, Processes, and Clinical Applications. Jones & Bartlett Publishers, pp. 35–53.

Bauer, A.J., Just, M.A., 2015. Monitoring the Growth of the Neural Representations of New Animal Concepts. Hum. Brain Mapp. 36, 3213–3226. https://doi.org/10.1002/hbm.22842

Bengio, Y., 2009. Learning Deep Architectures for AI. Found. Trends® Mach. Learn. 2, 1–127.

Bressler, S., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Brewer, G.J., Boehler, M.D., Pearson, R.A., DeMaris, A.A., Ide, A.N., Wheeler, B.C., 2009. Neuron network activity scales exponentially with synapse density. J. Neural Eng. 6, 014001. https://doi.org/10.1088/1741-2560/6/1/014001

Carpenter, G.A., Grossberg, S., 1988. The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. IEEE Comput. Soc. 21, 77–88. https://doi.org/10.1109/2.33

Case, R., 1987. Neo-Piagetian Theory: Retrospect and Prospect. Int. J. Psychol. 22, 773–791. https://doi.org/10.1080/00207598708246802

Casey, B.J., Tottenham, N., Liston, C., Durston, S., 2005. Imaging the developing brain: what have we learned about cognitive development? Trends Cogn. Sci. 9, 104–110. https://doi.org/10.1016/j.tics.2005.01.011

Cassimatis, N.L., 2012. Artificial Intelligence and Cognitive Modeling Have the Same Problem, in: Theoretical Foundations of Artificial General Intelligence. Atlantis Press, pp. 11–24.

Cho, S., Metcalfe, A., Young, C., Ryali, S., Geary, D., Menon, V., 2012. Hippocampal–Prefrontal Engagement and Dynamic Causal Interactions in the Maturation of Children's Fact Retrieval. J. Cogn. Neurosci. 24, 1849–1866. https://doi.org/10.1162/jocn_a_00246

Christensson, P., 2014. Flexible. TechTerms Dict.

Coenen, A., Zayachkivska, O., 2013. Adolf Beck: A pioneer in electroencephalography in between Richard Caton and Hans Berger. Adv. Cogn. Psychol. 9, 216–221. https://doi.org/10.2478/v10053-008-0148-3

Commons, M.L., 2008. Stacked Neural Networks must emulate evoltuion's hierarchical complexity. World Futur. 64, 444–451. https://doi.org/10.1080/02604020802301568

Commons, M.L., Gane-McCalla, R., Barker, C.D., Li, E., 2014a. The model of hierarchical complexity as a measurement system. Behav. Dev. Bull. 19, 9–14. https://doi.org/10.1037/h0100583

Commons, M.L., Goodheart, E., Pekker, A., Dawson-Tunik, T., Draney, K., Adams, K., 2008. Using Rasch Scaled Stage Scores to Validate Orders of Hierarchical Complexity of Balance Beam Task Sequences. J. Appl. Meas. 9, 182–199.

Commons, M.L., Goodheart, E., Richards, F., 1995. Primary, Concrete, Abstract, Formal, Systematic, and Metasystematic Balance Beam Problems. 1995 Pilot Versions.

Commons, M.L., Li, E., Richardson, A.M., Gane-McCalla, R., Barker, C.D., Tuladhar, C., 2014b. Does the model of hierarchical complexity produce significant gaps between orders and are the orders equally spaced? J. Appl. Meas. 15, 422–49.

Commons, M.L., Liu, X., 2017. Value, Delay Discounting and Risk as Seen at Three Levels of Analysis.

Commons, M.L., Miller, L., Giri, S., 2014c. A model of stage change explains the average rate of stage of development and its relationship to the predicted average stage (“smarts”). Behav. Dev. Bull. 19, 1–11. https://doi.org/10.1037/h0101076

Commons, M.L., Miller, P.-M., 2011. Folk psychology and criminal law: Why we need to replace folk psychology with behavioral science. J. Psychiatry Law 39, 493–516.

Commons, M.L., Pekker, A., 2008. Hierarchical Complexity: A Formal Theory. World Futur. 64, 375–382. https://doi.org/10.1080/02604020802301204

Commons, M.L., Robinett, T.L., 2013. Adult Development: Predicting Learning Success. Train. Ind. Q. Spring, 28–31.

Damon, F., Quinn, P.C., Heron-Delaney, M., Lee, K., Pascallis, O., 2016. Development of category formation for faces differing by age in 9- to 12-month-olds: An effect of experience with infant faces. Br. J. Dev. Psychol. 34, 582–597. https://doi.org/10.1111/bjdp.12152

Dandurand, F., Shultz, T., 2009. Modeling the Acquisition of a Torque Rule on the Balance scale Task, in: Proceedings of the 31st Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin, TX, pp. 1541–1546.

Dasaratha, S., 1996. Process modeling using stacked neural networks. Iowa State University.

Dawson, M., Zimmerman, C., 2003. Interpreting the Internal Structure of a Connectionist Model of the Balance Scale Task. Brain Mind 4, 129–149.

Dawson, T., Xie, Y., Wilson, M., 2003. Domain-general and domain-specific developmental assessments: do they measure the same thing? Cogn. Dev. 18, 61–78. https://doi.org/10.1016/S0885-2014(02)00162-4

Dawson-Tunik, T., Commons, M., Wilson, M., Fischer, K., 2005. The shape of development. Eur. J. Dev. Psychol. 2, 163–195. https://doi.org/10.1080/17405620544000011

Dawson-Tunik, T., Goodheart, E., Draney, K., Commons, M.L., 2010. Concrete, Abstract, Formal, and Systematic Operations as Observed in a “Piagetian” Balance-Beam Task Series. J. Appl. Meas. 11, 11–23.

De Beule, J., Bergen, B., 2006. On the emergence of compositionality, in: Proceedings of the 6th International Conference on the Evolution of Language. Presented at the 6th International Conference on the Evolution of Language, Rome, Italy. https://doi.org/10.1142/9789812774262_0005

Demetriou, A., Valanides, N., 1998. A three-level theory of the developing mind: Basic principles and implications for instruction and assessment., in: Intelligence, Instruction, and Assessment: Theory into Practice. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 149–199.

Dong, L., Wei, F., Zhou, M., Xu, K., 2014. Adaptive Multi-Compositionality for Recursive Neural Models with Applications to Sentiment Analysis, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Presented at the AAAI Conference on Artificial Intelligence, Québec, Canada, pp. 1537–1543.

Dueker, G., Needham, A., 2005. Infants’ object category formation and use: Real-world context effects on category use in object processing. Vis. Cogn. 12, 1177–1198. https://doi.org/10.1080/13506280444000706

Duran, R., Sorva, J., Leite, S., 2018. Towards an analysis of programs complexity from a cognitive perspective, in: Proceedings of the 2018 ACM Conference on International Computing Education Research. Presented at the International Computing Education Research Conference, Espoo, pp. 21–30. https://doi.org/10.1145/3105726.3105731

Elman, J., 1993. Learning and development in neural networks: the importance of starting small. Cognition 48, 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., 2010. Why Does Unsupervised Pre-training Help Deep Learning? J. Mach. Learn. Res. 11, 625–660. https://doi.org/10.1145/1756006.1756025

Fahlman, S.E., Lebiere, C., 1990. The Cascade Correlation Learning Architecture, in: Advances in Neural Information Processing Systems 2. Wiley-Interscience, Hoboken, pp. 524–532.

Farroni, T., Johnson, M.H., Menon, E., Zulian, L., Faraguna, D., Csibra, G., 2005. Newborns' preference for face-relevant stimuli: Effects of contrast polarity. Proc. Natl. Acad. Sci. 102, 17245–17250. https://doi.org/10.1073 pnas.0502205102

Featherson, K.G., Chen, S.J., Toth-Gauthier, M., Day, J.M., Philippe, H., Laverdeur, R., Nicolai, L., Nicolay, A.-C., Sini, A., Tilkin, C., Tyberghein, M., Vanheck, C., 2016. An Analysis of the Verbal Comprehension Index of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) Using the Model of Hierarchical Complexity (MHC): Why Might Stage Be a Better Measure of "Smarts" Than Verbal IQ? Behav. Dev. Bull. 21, 50–62. https://doi.org/10.1037/bdb0000015

Fernandes, F., Lona, L., 2005. Neural network Applications in Polymerization Processes. Braz. J. Chem. Eng. 22, 401–418. https://doi.org/10.1590/S0104-66322005000300009

Ferretti, R.P., Butterfield, E.C., 1986. Are children's rule assessment classifications invariant across instances of problem types? Child Dev. 57, 1419–1428. https://doi.org/10.2307/1130420

Fischer, K., 2008. Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education, in: The Educated Brain. Cambridge University Press, Cambridge, UK, pp. 127–150.

Fischer, K., 1987. Relations between Brain and Cognitive Development. Child Dev. 58, 623–632. https://doi.org/10.2307/1130202

Fischer, K., 1980. A theory of cognitive development: The control and construction of hierarchies of skills. Psychol. Rev. 87, 477–531. https://doi.org/10.1037/0033-295X.87.6.477

Flexibility, 2010. . Oxf. Dict. Engl.

French, R.M., 1999. Catastrophic Forgetting in Connectionist Networks: Causes, Consequences and Solutions. Trends Cogn. Sci. 3, 128–135. https://doi.org/10.1016/S1364-6613(99)01294-2

Fritsch, S., Guenter, F., Sulling, M., Mueller, S., 2016. Package "neuralnet", version 1.33: Training of Neural Networks.

Furtuna, R., Curteanu, S., Leon, F., 2012. Multi-objective Optimization of a Stacked Neural Network Using an Evolutionary Hyper-Heuristic.

Appl. Soft Comput. 12, 133–144. https://doi.org/10.1016/j.asoc.2011.09.001

Geary, D., 2011. Cognitive Predictors of Achievement Growth in Mathematics: A 5-Year Longitudinal Study. Dev. Psychol. 47, 1539–1552. https://doi.org/10.1037/a0025510

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural Networks and the Bias/Variance Dilemma. Neural Comput. 4, 1–58. https://doi.org/10.1162/neco.1992.4.1.1

Ghorbani, A., Owrangh, K., 2001. Stacked generalization in neural networks: generalization on statistically neutral problems, in: Proceedings of the International Joint Conference on Neural Networks. Washington. https://doi.org/10.1109/IJCNN.2001.938420

Giri, S., Commons, M.L., Harringan, W.J., 2014. There is only one stage domain. Behav. Dev. Bull. 19, 51–61. https://doi.org/10.1037/h0101081

Goodfellow, I.J., Mirza, M., Xiao, D., Courville, A., Bengio, Y., 2015. An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks. Cornell Univ. Libr.

Harringan, W.J., Commons, M.L., 2014. Stage of development of a species predicts the number of neurons. Behav. Dev. Bull. 19, 12–21. https://doi.org/10.1037/h0101077

Herculano-Houzel, S., 2012. How many neurons make a human brain? Billions fewer than we thought. Guard. - Neurosci.

Herculano-Houzel, S., 2009. The Human Brain in Numbers: A Linearly Scaled-up Primate Brain. Front. Hum. Neurosci. 3. https://doi.org/10.3389/neuro.09.031.2009]

Herd, S.A., O'Reilly, R., Hazy, T., Chattham, C., Brant, A.M., Friedman, N.P., 2014. A Neural Network Model of Individual Differences in Task Switching Abilities. Neuropsychologia 62, 375–389. https://doi.org/10.1016/j.neuropsychologia.2014.04.014

Hinton, G., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hinton, G., Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507. https://doi.org/10.1126/science.1127647

Hudspeth, W., Pribram, K., 1990. Stages of Brain and Cognitive Maturation. J. Educ. Psychol. 82, 881–884. https://doi.org/10.1037/0022-0663.82.4.881

Hume, D., 1888. A treatise of Human Nature, 2nd ed. Oxford University Press.
Husserl, E., 1999. The Idea of Phenomenology, 1999th ed. Springer, Netherlands.
Igel, C., Husken, M., 2003. Empirical evaluation of the improved Rprop learning algorithms. Neurocomputing 50, 105–123. https://doi.org/10.1016/S0925-2312(01)00700-7
Inhelder, B., Piaget, J., 1958. The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures. Routledge, London.
Irsoy, O., Cardie, C., 2014. Deep recursive neural networks for compositionality in language, in: Proceedings of the 27th International Conference on Neural Information Processing Systems. Presented at the NIPS'14, MIT Press Cambridge, Montreal, Canada, pp. 2096–2104.
Jasen, B.R.J., Maas, H.L.J. van der, 2001. Evidence for the Phase Transition from Rule I to Rule II on the Balance Scale Task. Dev. Rev. 21, 450–494. https://doi.org/10.1006/drev.2001.0530
Jasen, B.R.J., Maas, H.L.J. van der, 1997. Statistical Test of the Rule Assessment Methodology by Latent Class Analysis. Dev. Rev. 17, 321–357. https://doi.org/10.1006/drev.1997.0437
Johnson, M., 2011. Interactive Specialization: A domain-general framework for human functional brain development? Dev. Cogn. Neurosci. 1, 7–21. https://doi.org/10.1016/j.dcn.2010.07.003
Kanwisher, N., 2010. Functional specificity in the human brain: A window into the functional architecture of the mind. Proc. Natl. Acad. Sci. 107, 11163–11170. https://doi.org/10. 1073/pnas.1005062107/-/DCSupplemental.
Kashdan, T.B., Rottenberg, J., 2010. Psychological flexibility as a fundamental aspect of health. Clin. Psychol. Rev. 30, 865–878. https://doi.org/10.1016/j.cpr.2010.03.001
Klahr, D., Siegler, R., 1978. The representation of children's knowledge. Adv. Child Dev. Behav. 12, 61–116.
Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29, 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3
Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 4, 1–28. https://doi.org/10.3389/neuro.06.004.2008
Leber, A.B., Turk-Browne, N.B., Chun, M.M., 2008. Neural predictors of moment-to-moment fluctuations in cognitive flexibility. Proc. Natl.

Acad. Sci. 105, 13592–13597. https://doi.org/10.1073 pnas.0805423105

Lee, K.-H., Choi, Y.-Y., Gray, J., Cho, S.-H., Chae, J.-H., Lee, S., Kim, K., 2006. Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. NeuroImage 29, 578–586. https://doi.org/10.1016/j.neuroimage.2005.07.036

Leite, S., 2016. Successes in Cultural Evolution Raises the Variability in Humans' Highest Stage Attained. Behav. Dev. Bull. 21, 165–175. https://doi.org/10.1037/bdb0000033

Leite, S., Barker, C.D., Lucas, M.G., 2016. Neural Correlates of Postformal Stages of Reasoning: Biological Determinants of Developmental Stage. Behav. Dev. Bull. 21, 33–43. https://doi.org/10.1037/bdb0000012

Leite, S., Príncipe, J.-C., Ramos-Silva, A., Miller, P.-M., Marques-Teixeira, J., Commons, M.L., Rodrigues, P., submitted. Connectionist models capture hierarchical complexity transitions in development: discrimination between memory-based and operationally-based transitions. Cognition.

Leite, S., Ramos-Silva, A., Miller, P.-M., Commons, M.L., Rodrigues, P., 2018. Simulating Developmental Cognition: Learning by Order of Complexity in Modular Neural Networks. Prep.

Leite, S., Rodrigues, P., 2018. Simulating Developmental Cognition: Learning by Order of Complexity in Modular Stacked Neural Networks. MathPsych Conference, Wisconsin.

Leon, F., Zaharia, M., 2010. Stacked Heterogeneous Neural Networks for Time Series Forecasting. Math. Probl. Eng. 2010, 1–20. https://doi.org/10.1155/2010/373648

Maas, H.L.J. van der, Quinlan, P., Jansen, B., 2007. Towards better computational models of the balance scale task : A reply to Shultz and Takane. Cognition 103, 473–479. https://doi.org/10.1016/j.cognition.2007.01.009

Magalhães, L., 2014. Airport Flexibility: an Extensive Literature on its Origins, Definitions, Frameworks, Measurement and Case Studies (Mid-Term Report on Doctoral Study). Instituto Superior Técnico Universidade de Lisboa.

Maturana, H., Varela, F., 1928. Autopoesis and Cognition: The Realization of the Living. D. Reidel Publishing Company, Boston.

McClelland, J., 1995. A Connectionist Approach to Knowledge and Development, in: Developing Cognitive Competence: New Approaches to Process Modeling. LEA, Mahwah, NJ, pp. 157–204.

McClelland, J., 1989. Parellel distributed processing: implications for cognition and development.

McClelland, J., McNaughton, B., O'Reilly, R., 1995. Why there are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. Psychol. Rev. 102, 419–457. https://doi.org/10.1111/j.1551-6709.2011.01214.x

McCloskey, M., Cohen, N.J., 1989. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychol. Learn. Motiv. Adv. Res. Theory 24, 109–165. https://doi.org/10.1016/S0079-7421(08)60536-8

McFadden, G.T., Dufresne, A., Kobasigawa, A., 1987. Young children's knowledge of balance scale problems. Res. Theory Hum. Dev. 148, 79–94. https://doi.org/10.1080/00221325.1987.9914539

Mengistu, H., Huizinga, J., Mouret, J.B., Clune, J., 2016. Evolutionary Origins of Hierarchy. PLOS Comput. Biol. 12, 1–23. https://doi.org/10.1371/journal.pcbi.1004829

Mitchell, M., 1998. A complex-systems perspective on the "computation vs. dynamics" debate in cognitive science, in: 20th Annual Conference of the Cognitive Science Society.

Morita, T., Minoru, A., Naito, E., 2016. Contribution of Neuroimaging Studies to Understanding Development of Human Cognitive Brain Functions. Front. Hum. Neurosci. 10, 464–478. https://doi.org/10.3389/fnhum.2016.00464

Morris, M., 2006. An Introduction to the Philosophy of Language. Cambridge University Press.

Nath, S., 2013. Behaviourism as a Precursor of Identity Theory of Mind. Int. J. Sci. Res. Publ. 3, 1–7.

Nguyen, T., 2010. Total Number of Synapses in the Adult Human Neocortex. Undergrad. J. Math. Model. One Two 3, 1–14. https://doi.org/10.5038/2326-3652.3.1.26

Norris, D., 1990. How to build a connectionist idiot savant. Cognition 35, 277–291. https://doi.org/10.1016/0010-0277(90)90025-F

Pennachin, C., Goertzel, B., 2007. Contemporary Approaches to Artificial General Intelligence, in: Artificial General Intelligence. Springer-Verlag, Berlin, Heidelberg.

Peters, L., Smedt, B., 2018. Arithmetic in the developing brain: A review of brain imaging studies. Dev. Cogn. Neurosci. 30, 265–279. https://doi.org/10.1016/j.dcn.2017.05.002

Piuleac, C., Poulios, I., Leon, F., Curteanu, S., Kouras, A., 2010. Modeling Methodology Based on Stacked Neural Networks Applied to the

Photocatalytic Degradation of Triclopyr. Sep. Sci. Technol. 45, 1644–1650. https://doi.org/10.1080/01496395.2010.487736

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D., Menon, V., 2014. Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nature 17, 1263–9. https://doi.org/10.1038/nn.3788

Reyes, G., Osorio, O., Amy, B., 1997. Neural Learning of "High Level Rules": The Balance Scale Problem. Presented at the International Workshop on Neural Networks, Montreux, Swiss.

Richardson, F., Baughman, F., Forrester, N., Thomas, M., 2018. Computational Modeling of Variability in the Balance Scale Task.

Rijn, H., Someren, M. van, Maas, H.L.J. van der, 2003. Modeling developmental transitions on the balance scale task. Cogn. Sci. 27, 227–257. https://doi.org/10.1207/s15516709cog2702_4

Rojas, R., 1996. Neural Networks: A Systematic Introduction, Springer-Verlag. ed. Berlin.

Ross, A.M., Rhodes, D.H., Hastings, D.E., 2008. Defining Changeability: Reconciling Flexibility, Adaptability, Scalability, Modifiability, and Robustness for Maintaining System Lifecycle Value. Syst. Eng. 11, 246–262. https://doi.org/10.1002/sys.20098

Sage, S., Langley, P., 1983. Modeling cognitive development on the balance scale task. IJCAI83 Proc. Eighth Int. Jt. Conf. Artif. Intell. 1, 94–96.

Schapiro, A., McClelland, J., 2009. A Connectionist Model of a Continuous Developmental Transition in the Balance Scale Task. Cognition 110, 395–411. https://doi.org/10.1016/j.cognition.2008.11.017

Schulz, A.P., Fricke, E., Igenbergs, E., 2000. Enabling Changes in Systems throughout the Entire Life-Cycle – Key to Success?, in: Proceedings of the 10th Annual INCOSE Conference. Presented at the INCOSE Conference, Minneapolis, USA, pp. 565–573. https://doi.org/10.1002/j.2334-5837.2000.tb00426.x

Scott, W.A., 1962. Cognitive Complexity and Cognitive Flexibility. Sociometry 25, 405–414. https://doi.org/10.2307/2785779

Seipone, T., Bullinaria, J.A., 2005. The Evolution of Minimal Catastrophic Forgetting in Neural Systems, in: Proceedings of the 27th Annual Conference of the Cognitive Science Society. Presented at the XXVII Annual Conference of the Cognitive Science Society, Cognitive Science Society, Stresa, Italy, pp. 1991–1996.

Shultz, T., Cohen, L., 2004. Modeling Age Differences in Infant Category Learning. Infancy 5, 153–171. https://doi.org/10.1207/s15327078in0502_3

Shultz, T., Mareschal, D., Schmidt, W., 1994. Modeling Cognitive Development on Balance Scale Phenomena. Mach. Learn. 16, 57–86. https://doi.org/10.1023/A:1022630902151

Shultz, T., Schmidt, W., 1991. A Cascade-Correlation Model of the Balance Scale Phenomena. Proc. Thirteen. Annu. Conf. Cogn. Sci. Soc. 635–640.

Shultz, T., Schmidt, W., Buckingham, D., Mareschal, D., 1995. Modeling cognitive development with a generative connectionist algorithm, in: Developing Cognitive Competence: New Approaches to Process Modeling. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, US, pp. 205–261.

Siegler, R., Chen, Z., 2002. Development of Rules and Strategies: Balancing the Old and the New. J. Exp. Child Psychol. 81, 446–457. https://doi.org/10.1006/jecp.2002.2666

Sloutsky, V., 2010. From Perceptual Categories to Concepts: What Develops? Cogn. Sci. 34, 1244–86. https://doi.org/10.1111/j.1551-6709.2010.01129.x.

Smith, L., 2005. Cognition as a dynamic system: Principles from embodiement. Dev. Rev. 25, 278298. https://doi.org/10.1016/j.dr.2005.11.001

Smith, L., Sheya, A., 2010. Is Cognition Enough to Explain Cognitive Development? Top. Cogn. Sci. 2, 725–735. https://doi.org/10.1111/j.1756-8765.2010.01091.x

Spencer, J., Austin, A., Schutte, A., 2012. Contributions of Dynamic Systems Theory to Cogntive Development. Cogn. Dev. 27, 401–418. https://doi.org/10.1016/j.cogdev.2012.07.006

Stability, 2010. . Oxf. Dict. Engl.

Supekar, K., Swigart, A., Tenison, C., Jolles, D., Rosenberg-Lee, M., Fuchs, L., Menon, V., 2013. Neural predictors of individual differences in response to math tutoring in primary-grade school children. Proc. Natl. Acad. Sci. 110, 8230–8235. https://doi.org/10.1073/pnas.1222154110

Sweller, J., 2004. Instructional Design Consequences of an Analogy between Evolution by Natural Selection and Human Cognitive Architecture. Instr. Sci. 32, 9–31. https://doi.org/10.1023/B:TRUC.0000021808.72598.4d

Taylor, P., Hobbs, J.N., Burroni, J., Siegelmann, H.T., 2015. The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. Nat. Sci. Rep. 5, 18112–18130.

Ting, K., M., Witten, I.H., 1999. Issues in Stacked Generalization. J. Artif. Intell. Researc 10, 271–289. https://doi.org/10.1613/jair.594

Tudor, M., Car, L., Tudor, K.I., 2005. Hans Berger (1873-1941)--the history of electroencephalography. Acta Media Croat. 59, 307–313.

Wang, L.-C., Der, S.Z., Nasrabadi, N., 1998. Automatic target recognition using a feature-decomposition and data-decomposition modular neural network. IEEE Trans. Image Process. 7, 1113–1121. https://doi.org/10.1109/83.704305

Wendelken, C., Ferrer, E., Whitaker, K., Bunge, S., 2015. Fronto-Parietal Network Reconfiguration Supports the Development of Reasoning Ability. Cereb. Cortex 26, 2178–2190. https://doi.org/10.1093/cercor/bhv050

Wilkening, F., Anderson, N.H., 1982. Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structures. Psychol. Bull. 92, 215–237. https://doi.org/10.1037/0033-2909.92.1.215

Zhang, J., Martin, E., Morris, J., Kiparissides, C., 1997. Inferential Estimation of Polymer Quality Using Stacked Neural Networks. Comput. Chem. Eng. 21, S1025–S1030. https://doi.org/10.1016/S0098-1354(97)87637-5

Zhou, Z.-H., Wu, J., Tang, W., 2002. Ensembling neural networks: Many could be better than all. Artif. Intell. 137, 239–263. https://doi.org/10.1016/S0004-3702(02)00190-X

Zimmerman, C., 1999. A network interpretation approach to balance scale tasks. University of Alberta, Canada.

Zimmerman, C., Croker, S., 2014. Cognitive Dynamics on the Balance-Scale Task. Proc. Annu. Meet. Cogn. Sci. Soc. 36, 3161–66.