

2014

Modelling Learning to Count in Humanoid Robots

Rucinski, Marek

<http://hdl.handle.net/10026.1/2995>

<http://dx.doi.org/10.24382/4503>

Plymouth University

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

Modelling Learning to Count in Humanoid Robots

by

Marek Ruciński

A thesis submitted to the Plymouth University
in partial fulfilment for the degree of

Doctor of Philosophy

School of Computing and Mathematics
Faculty of Science and Technology

March 2014

Marek Ruciński

Modelling Learning to Count in Humanoid Robots

Abstract

This thesis concerns the formulation of novel developmental robotics models of embodied phenomena in number learning. Learning to count is believed to be of paramount importance for the acquisition of the remarkable fluency with which humans are able to manipulate numbers and other abstract concepts derived from them later in life. The ever-increasing amount of evidence for the embodied nature of human mathematical thinking suggests that the investigation of numerical cognition with the use of robotic cognitive models has a high potential of contributing toward the better understanding of the involved mechanisms. This thesis focuses on two particular groups of embodied effects tightly linked with learning to count. The first considered phenomenon is the contribution of the counting gestures to the counting accuracy of young children during the period of their acquisition of the skill. The second phenomenon, which arises over a longer time scale, is the human tendency to internally associate numbers with space that results, among others, in the widely-studied SNARC effect. The PhD research contributes to the knowledge in the subject by formulating novel neuro-robotic cognitive models of these phenomena, and by employing these in two series of simulation experiments. In the context of the counting gestures the simulations provide evidence for the importance of learning the number words prior to learning to count, for the usefulness of the proprioceptive information connected with gestures to improving counting accuracy, and for the significance of the spatial correspondence between the indicative acts and the objects being enumerated. In the context of the model of spatial-numerical associations the simulations demonstrate for the first time that these may arise as a consequence of the consistent spatial biases present when children are learning to count. Finally, based on the experience gathered throughout both modelling experiments, specific guidelines concerning future efforts in the application of robotic modelling in mathematical cognition are formulated.

Contents

Acknowledgements	1
Author’s Declaration	3
1 Introduction	5
1.1 Motivation	5
1.2 Objectives, Scope, and Contribution of the Thesis	7
1.3 Structure of the Thesis	11
I Background	13
2 Numerical Knowledge in Humans	15
2.1 Elementary Human Numerical Abilities	16
2.1.1 Non-verbal Magnitude Processing	16
2.1.2 Counting	20
2.2 Embodied Facets of Human Number Knowledge	24
2.2.1 The Origins of Mathematical Ideas	25
2.2.2 The Involvement of Hands in Counting	27
2.2.3 The Association of Numbers and Space	31
2.3 The Development of Human Numerical Skills	40
2.3.1 The Development of Non-verbal Magnitude Processing	40
2.3.2 The Development of Counting	43
2.3.3 The Development of Spatial-Numerical Associations	54
2.4 Numerical Knowledge in Humans — Summary	60
3 Computational Modelling in Mathematical Cognition	65
3.1 Qualitative Models of Human Mathematical Skills	66
3.2 Models of Elementary Numerical Abilities	69
3.2.1 The First Quantitative Model	69
3.2.2 The Mental Representation of Magnitude	70
3.2.3 Between Subitising and Counting	73
3.2.4 The Temporal Structure of Numerical Processing	77
3.2.5 Identifying the Sources of the Behavioural Effects	79
3.2.6 Grounding Quantification in Perception	81
3.3 Models of Counting	82
3.3.1 Counting Identical Sequential Stimuli	82
3.3.2 Learning the Sequence of Number Words	83

3.3.3	Artificial Neural Network Architectures for Counting	84
3.3.4	Spontaneous Counting in an Elman Network	85
3.3.5	Multi-Net Simulation of Counting	87
3.4	Models of the SNARC Effect	91
3.5	Research on Gestures in Humanoid Robots	96
3.6	Summary	100
4	Methods	103
4.1	Embodied Cognition	103
4.2	Developmental Cognitive Robotics	106
4.3	iCub — Humanoid Robotic Platform	108
4.4	Artificial Neural Network Models Employed in the Study	112
4.4.1	Simple Recurrent Artificial Neural Networks	112
4.4.2	Continuous-time Neural Networks	115
4.4.3	Self-Organising Maps	117
4.5	Empirical Investigation of the Role of Gestures in Learning to Count	122
4.5.1	Experiment Design	123
4.5.2	Counting Error Types	125
4.5.3	Results	125
4.6	Plan of Work	127
II	Neuro-Robotic Model of Learning to Count	131
5	Model Overview	133
5.1	Model Design Assumptions	133
5.2	Model Architecture	135
5.2.1	Vision	135
5.2.2	Gestures	138
5.2.3	Speech and Number Words	142
5.3	Training of the Model	146
5.3.1	Preliminary Skill — Learning the Count List	147
5.3.2	Counting	149
5.3.2.1	Spatio-temporal Gestures	150
5.3.2.2	Rhythmic Gestures	152
5.4	Model Performance Evaluation	154
5.5	Model Implementation	159
5.6	Discussion	159
6	Simulations	163
6.1	Simulation 1 — Learning Number Words	163
6.1.1	Aims of the Experiment	163
6.1.2	Procedure	164
6.1.3	Results	165
6.1.4	Discussion	167
6.2	Simulation 2 — Preliminary Training Stage and Generalisation	172
6.2.1	Aims of the Experiment	172
6.2.2	Procedure	173
6.2.3	Results	174

6.2.4	Discussion	176
6.3	Simulation 3 — Contribution of the Counting Gestures	178
6.3.1	Aims of the Experiment	178
6.3.2	Procedure	178
6.3.3	Results	183
6.3.4	Discussion	184
6.4	Simulation 4 — Spatial Aspect of the Counting Gestures	190
6.4.1	Aims of the Experiment	190
6.4.2	Procedure	191
6.4.3	Results	193
6.4.4	Discussion	193
6.5	Summary	197

III Neuro-Robotic Model of the Acquisition of Spatial-Numerical Associations 199

7	Model Overview	201
7.1	Model Design Assumptions	202
7.2	Model Architecture	203
7.2.1	‘What’ Pathway	204
7.2.2	‘Where’ Pathway	206
7.2.3	Model Implementation	209
7.3	Developmental Learning	210
7.3.1	Building Spatial Representations and Transformations	211
7.3.2	Learning the Semantics of Number Symbols	213
7.3.3	Learning to Count	215
7.3.4	Learning Simple Numerical Tasks	216
7.4	Model Evaluation	218
7.5	Summary	219
8	Simulations	221
8.1	Simulation 1 — Results of the Development Process	221
8.1.1	Aims of the Experiment	221
8.1.2	Procedure	222
8.1.3	Results	222
8.1.4	Discussion	228
8.2	Simulation 2 — Number Size and Numerical Distance Effects	233
8.2.1	Aims of the Experiment	233
8.2.2	Procedure	233
8.2.3	Results	234
8.2.4	Discussion	234
8.3	Simulation 3 — The SNARC Effect	236
8.3.1	Aims of the Experiment	236
8.3.2	Procedure	236
8.3.3	Results	238
8.3.4	Discussion	238
8.4	Simulation 4 — The Posner-SNARC Effect	241
8.4.1	Aims of the Experiment	241

8.4.2	Procedure	241
8.4.3	Results	242
8.4.4	Discussion	244
8.5	Summary	244
9	Conclusions	247
9.1	Future Work — Research Questions 1–3	252
9.1.1	Comparison of the Experimental Protocols	253
9.1.2	Research Question 1	256
9.1.3	Research Questions 2 and 3	258
9.2	Future Work — Research Question 4	262
9.3	Future Prospects of Developmental Cognitive Robotics in Mathematical Cognition	267
A	Equations Describing the Spatial-Numerical Associations Model	273
A.1	Notation	273
A.2	Number Comparison Task	274
A.3	Parity Task	277
A.4	Visual Target Detection Task	278
	References	281
	Publications	297

List of Figures

1	Numerical knowledge development timeline	61
2	Architecture of the model of Chen and Verguts (2010)	95
3	iCub humanoid robot	110
4	Elman and Jordan networks	113
5	Self-organising maps	120
6	Experimental design of Alibali and DiRusso (1999)	124
7	Coding categories for children’s counting performance	126
8	Architecture of the model of learning to count with gestures	137
9	Visual input representation example	137
10	Counting gesture production using the iCub robot	143
11	Counting gesture joint angles	144
12	Architecture of the neural network in the first stage of the training.	148
13	Spatio-temporal counting gesture construction example	151
14	Rhythmic counting gesture construction example	153
15	Experimental design of a neuro-robotic simulation of counting	155
16	Correct model output example	157
17	Incorrect model output example	158
18	Number words learning progress in trial 028	166
19	Number words learning progress in trial 042	168
20	Number words learning progress in trial 075	169
21	Progress of number words learning across trials	170
22	Configuration of the model used in simulation 2	175
23	Profile plots for the simulation 2 ANOVA	175
24	Configurations of the model used in simulation 3	180
25	Counting accuracy of the model in simulation 3	185
26	Profile plot for the simulation 3 ANOVA	185
27	Errors made in child conditions and puppet conditions	186
28	Configurations of the model used in simulation 4	194
29	Profile plot for the simulation 4 ANOVA	194
30	Architecture of the model of spatial-numerical associations.	207
31	Connection weights between the input and the semantic layer	207
32	Development of the model of spatial-numerical associations	212

33	Motor babbling operational space	214
34	Motor babbling simulation	214
35	Response time measurement	218
36	Developed gaze SOM example	224
37	Developed left arm SOM example	225
38	Visualisation of the arm SOMs in the robot space	227
39	Connection weights in the Where pathway	229
40	Connection weights between the input layer and the gaze map	229
41	Results of the simple numerical tasks training.	230
42	Configuration of the model in the number comparison task	235
43	Number size and numerical distance effects simulation	235
44	Configuration of the model in the number parity task	239
45	SNARC effect simulation	239
46	Configuration of the model in the visual target detection task	243
47	Posner-SNARC effect simulation	243
48	Concept of the extension of the model	266

List of Tables

1	Preliminary training success rates	169
2	Counting errors made by the model	186
3	Output length regression analysis	186
4	SOM parameters	214
5	SOM quality across trials	227

Acknowledgements

The beginnings of my interest in the cognitive mechanisms that underlie human numerical reasoning can be traced as far back as the fall of 2002. Then, having had just commenced the BSc degree studies at the Faculty of Computer Science and Management (now Faculty of Computing) of the Poznań University of Technology, I attended, among others, a course in mathematics lead by Dr Jacek Gruszka. There are two things I vividly remember from these series of lectures, up to this very day. The first one is that the idempotents of an abelian monoid form its proper submonoid. The second one is Dr Gruszka's lecture about the set-theoretic definition of natural numbers. Therein, talented a speaker as he is, Dr Gruszka introduced the theory with the use of a thought experiment, in which the aim was to help understand what numbers are to an infant who does not yet have the ability to count. This was the first time I pondered the question of cognitive bases of mathematical knowledge, struck by the contrast between the apparent abstractness of the concept of number and the very down-to-earth character of the activity — comparing piles of candies — in which it could be grounded. The impression left by this lecture was strong enough that while travelling for an interview for the RobotDoC PhD studentship at Plymouth University approximately eight years later, I did not have the slightest doubts about what I am going to say when asked about a topic I would like to research with the use of developmental cognitive robotics.

Four years less one month from the date of this interview, I am at the point of submitting my PhD dissertation on modelling learning to count in robots and would like to use this opportunity to say thank you to everyone who was instrumental in me achieving this goal. First, there are several people my interaction with whom was an important factor in convincing me that scientific career is worth pursuing. Under the direction of Dr Grzegorz Pawlak of the Poznań University of Technology, who founded the Students' Association of Computing Sciences the same year I entered the University, I had an opportunity to make my first steps in science. Participation in the activities of the Association enabled me to conduct my first research projects, publish my first papers and attend my first scientific conferences, all this very early on into my academic education. Later on, it were Dr Leopold Summerer, Dr Dario Izzo, and Dr Christos Ampatzis, at the time all of the Advanced Concepts Team of the European Space Agency, who, during my traineeship in the team, reaffirmed me in the goal of pursuing the degree of the Doctor of Philosophy.

My PhD studentship at Plymouth University was an unforgettable experience, and to a large degree this has to be attributed to the people I had the pleasure to get to know and to work with during that time. In the first place, I would like to thank my supervisory team, that is Prof. Angelo Cangelosi, Prof. Tony Belpaeme and Dr Davide Marocco, for their scientific support. I am especially grateful to Prof. Cangelosi for his active interest in my work throughout the entire duration of my studentship, and for the time he has invested in guiding me through my PhD endeavour. I greatly benefited from his advice in the matters of scientific, and sometimes also non-scientific nature. Thank you as well to my 'extended supervisory team', which includes Dr Elena Dell'Aquila of the Plymouth University, for her great coaching work in the RobotDoC project, and Prof. Stefan Wermter of the University of Hamburg, for his hospitality, and his advice I benefited from during

my secondment in his lab.

I would like to thank my dear friends Francesca Stramandinoli and Federico Da Rold for being with me through the times good and bad, for all the lovely time we had in Plymouth (and outside of it), and for the sea of coffee we have drunk together.

Thanks also to Joachim de Greef and Robin Read, as well as to John, Zacch, and Ben Burrell for their efforts in organising basketball sessions I was participating in for nearly two years. It was invaluable to once in a while remind my brain, on a daily basis so occupied with science, that it is, first and foremost, embodied.

I would like to thank the European Commission Marie Curie Actions for their generous funding which not only allowed me to focus entirely on the scientific work, but also enabled me to participate fully in all aspects of contemporary academia.

I would like to say thank you also to Prof. Martin Fischer of the University of Potsdam, Prof. Tom Verguts of the Ghent University, and Prof. Marco Zorzi of the University of Padova, who, at various stages of my research, provided me with valuable suggestions which allowed me to significantly improve my work. Their feedback was so ample that, in fact, it proved to be impossible to incorporate all of it within the present thesis.

Last but not least, I would like to thank my beloved parents, Jolanta and Leszek Rucińscy, for their patience and forgiveness in face of the fact that during the final years of my degree I was not able to see them in person as often as I should and wished.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

This study was financed by the EU project RobotDoC (235065) from the Framework Programme 7, Marie Curie Actions Initial Training Network.

A programme of advanced study was undertaken, which included participation in eight biannual Training Milestones organised by the RobotDoC Collegium, an international visit at the University of Hamburg, Germany, participation in the relevant courses organised within the Researcher Development Programme of the Plymouth University Graduate School, and participation in the organisation of the Post-Graduate Conference on Robotics and Development of Cognition in the role of programme chair.

Relevant scientific seminars and conferences were regularly attended at which work was presented; external institutions were visited for consultation purposes and several papers prepared for publication.

Publications:

- Ruciński, M., Cangelosi, A. & Belpaeme, T. (2011). An embodied developmental robotic model of interactions between numbers and space. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *33rd annual meeting of the cognitive science society* (pp. 237–242).
- Stramandinoli, F., Ruciński, M., Znajdek, J., Rohlfing, K. J. & Cangelosi, A. (2011). From sensorimotor knowledge to abstract symbolic representations. *Procedia Computer Science*, 7, 269–271.
- Ruciński, M. (2011), Robotic models of mathematical cognition. In A. Wiśniewska et al. (Eds.), *Proceedings of the Science — Passion, Mission, Responsibilities — Marie Curie Researchers Symposium* (p. 61).
- Ruciński, M., Stramandinoli, F. (2012), An embodied view on the development of symbolic capabilities and abstract concepts. In J. Szufnarowska (Ed.), *Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition* (pp. 62–63).
- Ruciński, M., Cangelosi, A. & Belpaeme, T. (2012). Robotic model of the contribution of gesture to learning to count. In *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob 2012)* (pp. 1–6).

Attended Conferences and Summer Schools:

- Cognitive Robotics Research Methods Workshop (RobotDoC TM 1),
9–11/03/2010, Plymouth, UK
- Cognitive Science and Machine Learning Summer School (MLSS Sardinia 2010),
5–13/05/2010, Pula, Italy
- Marie Curie Conference (ESOF 2010 Satellite Event), 1–2/07/2010, Torino, Italy

Euroscience Open Forum (ESOF 2010), 2–7/07/2010, Torino, Italy

Project Proposal Workshop (RobotDoC TM 2), 25–27/10/2010, Bielefeld, Germany

BCS-SGAI 2010 Conference Workshop on Bio-inspired and Bio-Plausible Cognitive Robotics, 14/12/2010, Cambridge, UK

Spring School on Interdisciplinary Methods for Cognitive Robotics (RobotDoC TM 3), 2–4/05/2011, Budapest, Hungary

The European Future Technologies Conference and Exhibition *fet*¹¹, 4–6/05/2011, Budapest, Hungary

33rd Annual Conference of the Cognitive Science Society (CogSci 2011), 20–23/07/2011, Boston, MA, USA

PhD Transfer Workshop (RobotDoC TM 4), 5–7/09/2011, Barcelona, Spain

Science — Passion, Mission, Responsibilities – Marie Curie Researchers Symposium, 25–27/09/2011, Warsaw, Poland

5th International Conference on Cognitive Systems (CogSys 2012), 22–23/02/2012, Vienna, Austria

1st EUCogIII Members’ Conference ‘Do Robots Need Cognition? — Does Cognition need Robots?’, 23–24/02/2012, Vienna, Austria

RobotDoC Entrepreneurship Workshop (RobotDoC TM 5), 22–23/03/2012, Genova, Italy

10th Anniversary of European Space Agency’s Advanced Concepts Team Conference, 2–3/07/2012, Noordwijk, the Netherlands

Veni Vidi Vici 2012, the iCub Summer School, 18–27/07/2012, Sestri Levante, Italy

Postgraduate Conference on Robotics and Development of Cognition (RobotDoC TM 6), 10–12/09/2012, Lausanne, Switzerland

IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob 2012), 7–9/11/2012, San Diego, CA, USA

Spring School on Developmental Robotics and Cognitive Bootstrapping (RobotDoC TM 7), 18–20/03/2013, Athens, Greece

RobotDoC International Conference on Development of Cognition (RobotDoC TM 8), 16–18/08/2013, Osaka, Japan

External Contacts:

Prof. Stefan Wermter, Knowledge Technology Lab, University of Hamburg

Word count of main body of thesis: 71,660

Signed

Date

Chapter 1

Introduction

1.1 Motivation

The ability to perceive and process quantities and amounts appears to be one of the fundamental skills of animals (Dehaene, 1997). It is hard to argue with the observation that being able to choose a bigger pile of food, a larger prey, a tree that bears more fruit or a less densely populated area to inhabit, increases the chances of the creature's survival in the competitive natural environment. It is therefore not surprising that throughout a century of study of animal quantitative skills, numerical competence has been confirmed in a wide spectrum of species, at different levels of the biological complexity (Brannon, 2005; Pepperberg, 2006).

The quantification skills of animals cannot however be compared with, let alone match, those of humans. Developed and refined over the course of the last few thousand years, mathematics is regarded as deep, essential and fundamental to human experience (Lakoff & Núñez, 2000). Based on symbol manipulation and formal proofs, mathematics lies at the core of all other sciences. We use mathematical equations to describe the laws governing the world around us — be they physical, chemical, biological or sociological. In fact, being able to capture a phenomenon in an equation is often considered equivalent to actually *understanding* it. As mathematics can be put forward as a prime example of abstract thinking, it is not surprising

that it has been receiving a lot of attention from scientists who study human behaviour, the nature and structure of thought, and the inner workings of the brain. What makes it possible for the brain to construct the most fundamental abstract concepts, such as sets and numbers? How are these acquired in the course of the human cognitive development? How does the brain discover and formalise the rules that govern these concepts? What enables these ideas to be extended toward concepts that have so little connection with our material experience, such as negative numbers or algebras? What perceptual and cognitive mechanisms are necessary to achieve the required level of abstraction? Addressing these and many other related questions has been the subject of research that has been put together under the collective name *mathematical cognition* (Campbell, 2005).

The scientists who study mathematical cognition employ a broad range of means to find the answers to the questions they tackle. The research in experimental psychology provides both cross-sectional and longitudinal behavioural data about the quantitative skills of animals and humans. Anthropologists study how mathematical concepts and notations have been changing along with the expansion of the human civilisation, and together with the related aspects of human culture, such as language. Cognitive scientists investigate the nature of mathematical thought. Neuroscientists in turn, endeavour to pin down the neural circuitry that is involved in the representation and processing of numbers. Finally, the models of human numerical capabilities, formulated based on the data gathered by all the above, can be subjected to both qualitative and quantitative analysis with the use of computational modelling (Zorzi, Stoianov & Umiltà, 2005).

There is however another, relatively recently proposed tool to study cognition, which — to the best of my knowledge — up to this day has not yet been employed in the context of mathematical thought, namely developmental cognitive robotics (Asada, MacDorman, Ishiguro & Kuniyoshi, 2001). Originally proposed as a paradigm for the design of intelligent robots, it provides an appealing framework for the implementation and assessment of cognitive models. The main reason for

the attractiveness of the developmental cognitive robotics in the context of human mathematical thought lies in the fact that there is a growing amount of evidence for the highly embodied character of the latter (Lakoff & Núñez, 2000). Developmental cognitive robotics elegantly supplements the ‘traditional’ computational modelling methodology by taking into account an artificial model of the human body and, by extension, of the environment in which the body exists. In line with the embodied view of cognition, according to which the body and the learning environment play an important role in the emergence of intelligence, it can be argued that robotic modelling reduces the arbitrariness of the cognitive model, and increases its plausibility.

1.2 Objectives, Scope, and Contribution of the Thesis

In general terms, the main goal of the research described in this thesis is to apply developmental cognitive robotics in advancing the state of the art of the knowledge in certain areas of mathematical cognition. The focus is put on those aspects of human mathematical thought which have strong relevance to the notion of embodiment, as this is the area where robotic modelling can be expected to be the most advantageous. More specifically, I look at *learning to count*, which is a distinct milestone in the human cognitive development, believed to be of profound importance for the scaffolding of mathematical reasoning. According to one of the hypotheses, through the acquisition of counting we link our pre-verbal and approximate quantification skills (available also to lower animals) with our precise symbol manipulation capabilities (Le Corre & Carey, 2007). This enables us to employ the power of the latter in the context of operations on quantities, giving birth to mathematics. Due to its importance, learning to count has been receiving significant attention on the part of the experimental psychologists and of the computational cognitive modelling community for a considerable time now. Despite this, there are still many open ques-

tions about the acquisition of counting which require further investigation. Present work aims at answering the following four research questions:

Research Question 1. How does mastering the count list prior to learning to count within the respective range of collection sizes affect the subsequent process of learning to count?

Research Question 2. Can counting gestures, represented in the form of the values of arm joint angles that change over time, contribute toward the improvement of the counting accuracy?

Research Question 3. Is the spatial correspondence between the items being enumerated and the indicating act performed during counting an important characteristic of the counting gestures?

Research Question 4. Can consistent spatial biases present in children's learning to count be a source of the spatial-numerical associations later in life?

In the search for the answers to these questions, this thesis looks at two embodied phenomena connected with learning to count, which are associated with different developmental time scales. The first one, connected primarily with the research questions 2 and 3, is the contribution of the counting gestures to learning to count (Graham, 1999). Ample experimental evidence shows, that during the critical period of the development of the skill, the counting gestures, such as pointing to or touching the items one by one as they are enumerated, play an important facilitating role. This takes place over a relatively short period of no more than 4 years during infancy. Although several hypotheses concerning the phenomenon exist, as of today the exact mechanism of this contribution has not been established.

The second embodied effect connected with learning to count, considered herein in relation to the research question 4, are so-called spatial-numerical associations (Göbel, Shaki & Fischer, 2011). They are believed to emerge in the childhood and continue to be present in the adulthood. In general terms, they are evident in that humans tend to internally associate small numbers with the left side of space

and large numbers with the right side of space. One of the most-widely studied experimental manifestations of the spatial-numerical associations is the SNARC effect (Dehaene, Bossini & Giraux, 1993), a particular pattern found in the left- and right-handed response times in timed execution of simple numerical task. The significance of the investigation of the spatial-numerical associations is connected with the insight it is hoped to provide into the nature of the representation and processing of numbers and quantities by the brain. Their ontogeny — referred to in the research question 4 — is one of the questions about the spatial-numerical associations which still remain to be answered, despite the intensive study they have been subject to since their discovery.

As the result of this dual point of view on learning to count, the present thesis comprises two sets of robotic simulation experiments. Although distinct and associated with separate computational models, these experiments are unified by the central role of the same developmental phenomenon (i.e. learning to count) in the matter under investigation, and by the general framework of the model formulation and realisation (i.e. developmental cognitive robotics).

In the light of the available past work in the considered areas, the original contribution of the present PhD research can be summarised as follows:

- first extensive application of the developmental cognitive robotics methodology in the context of mathematical cognition;
- formulation of a novel, developmental neuro-robotic model of learning to count, aimed at the investigation of the contribution of the gestures to the acquisition of the counting skill;
- validation of the model in terms of its ability to generalise to novel input patterns;
- evidence found through the model simulations for the importance of the acquisition of the count list prior to the commencement of learning to count (research question 1);

- evidence found through the model simulations for the usefulness of the proprioceptive information conveyed by the counting gestures in the context of learning to count (research question 2);
- evidence found through the model simulations for the importance of the spatial correspondence between the counting gestures and the enumerated items (research question 3);
- formulation, based on the experience gathered throughout the implementation and simulation of the model, of specific guidelines for future research in cognitive modelling of learning to count;
- extension of an antecedently published model of spatial-numerical associations (Chen & Verguts, 2010) with richer representations of the relevant embodied aspects of the human sensorimotor development, and formulation, as a result, of a neuro-robotic model which focuses on the ontogeny of the spatial-numerical associations;
- validation of the model in terms of its ability to qualitatively reproduce a range of behavioural effects encountered in the context of the spatial-numerical associations, namely the number size effect, the numerical distance effect, the SNARC effect and the Posner-SNARC effect;
- demonstration through the model simulations that the effects connected with the spatial-numerical associations (such as the SNARC effect) may emerge as the result of the systematic spatial biases present when children learn to count (research question 4). This is a prediction resulting from the proposed model, which has found additional support in the experimental findings published in parallel (Opfer, Thompson & Furlong, 2010).
- as a consequence of the above, validation of the model of Chen and Verguts (2010) in terms of the plausibility of the assumptions made by its authors regarding the ontogeny of the spatial-numerical associations;

- formulation, based on the experience gathered throughout the implementation and simulation of the model, of specific guidelines for future research in cognitive modelling of spatial-numerical associations and their acquisition, including a concept of an extended model argued to be able to account, for the first time, for certain additional properties of the SNARC effect.

1.3 Structure of the Thesis

This thesis is divided logically into three parts. Part I, consisting of three chapters, introduces the appropriate background for the present study. Chapters 2 and 3 provide an overview of the relevant past research. Chapter 2 summarises what is known about the elementary human numerical skills, focusing primarily on the results of studies in experimental psychology. I introduce the most important findings about the behavioural phenomena that are at the centre of interest in this thesis, namely the ability to non-verbally process quantities, the ability to count, and the spatial-numerical associations. At the same time, the evidence is provided in support of the view that mathematical thinking in humans is, to a large extent, embodied. This constitutes the primary argument for introducing herein developmental cognitive robotics as an aid in its study. Chapter 3 establishes the context for the modelling experiments presented in this thesis, by reviewing the past efforts in the computational modelling of the aforementioned aspects of human numerical capabilities. The importance of this is based on two reasons. First, my modelling work builds on the already existing models, in the sense that past work is often being referred to, for example when design decisions are explained. Second, having a thorough review of the existing research at hand allows the novel contribution to knowledge of my work to be clearly highlighted. Finally, chapter 4 introduces the methodological aspects of the present neuro-robotics study. This starts with a brief presentation of the embodied view of cognition and of the developmental cognitive robotics paradigm. This chapter also deals with the issues related to the implement-

ation of the models presented in this thesis. I describe the humanoid robot that is used as the target platform for the experiments, discuss the artificial neural network frameworks, on which the implementation of my cognitive models is based, and also briefly review a particular experimental study, after which some of the simulations conducted herein have been modelled.

The second and third part of this thesis present, respectively, two neuro-robotic models of the embodied phenomena connected with learning to count. Both these parts consists of two chapters. The first chapter in each part presents the model itself, and the associated procedures connected with its training and evaluation. The second chapter in each part describes the simulations conducted using the model and their results. The model introduced in part II (chapters 5 and 6), aims to tackle the research questions 1–3, and looks at the contribution of the counting gestures to learning to count. The second model, which is the subject of part III of this thesis (chapters 7 and 8), considers a more extended developmental time scale and attempts to provide an answer to the research question 4. The thesis concludes with chapter 9, in which the obtained results are summarised and discussed.

Part I

Background

Chapter 2

Numerical Knowledge in Humans

In this chapter the current knowledge about the fundamental human numerical abilities is reviewed. Given that the range of topics in mathematical cognition is extremely broad (see Campbell, 2005), the discussion presented here is necessarily limited to focus only on those that are directly relevant to the present study. Unfortunately, this implies that several exciting themes, such as the research on quantification skills of animals or the anthropological findings about the number notation systems used by humans throughout the centuries, had to be omitted altogether.

This chapter starts (section 2.1) with an overview of the methods of investigation and of the basic findings about two principal numerical skills of humans — the capability to *non-verbally process numerosity*, and an activity in which it is believed to become linked with language, that is *counting*. This is followed by section 2.2 which presents the selected arguments that can be put forward in support of the hypothesis that the origins of the human numerical knowledge are embodied. This includes a discussion of the phenomena which are addressed by the four research questions considered in this thesis: section 2.2.2 discusses what is known about the *contribution of the counting gestures to learning to count*, while section 2.2.3 is an overview of the findings connected with *spatial-numerical associations*. Finally, section 2.3 provides in-depth information about the *developmental trajectories* of these two phenomena. The decision to review the research on the ontogeny separately

was motivated by its importance from the point of view of the developmental cognitive robotics methodology (cf. section 4.2), the principal tool chosen to tackle the research questions posed in this thesis.

2.1 Elementary Human Numerical Abilities

2.1.1 Non-verbal Magnitude Processing

Among the most fundamental numerical skills possessed by humans is the ability to perceive and process the numerosity of a set conveyed non-verbally, that is directly through its spatial or temporal characteristics. This skill has been shown to be available to some of the animal species as well, it is widely accepted to reflect our ability to represent mentally the magnitudes of numbers in an approximate, ‘analogue’ way, and is sometimes called ‘the number sense’ (Dehaene, 1997). It is manifested for instance in the ability to estimate the size of a set or to tell which of the two sets is larger, without counting. While numerosity estimation requires the possession of a symbolic representation of numbers of some sort (e.g. number words), the comparison does not; therefore, the latter is among the most ubiquitous tasks in the cognitive study of non-verbal mathematical abilities both in humans and in animals.

The magnitude processing by human adults is characterised by several robust effects, that can be observed through the analysis of the response times (RT) or of the error rates, and that provide insight into the nature of the number representation in the brain. The *numerical distance effect* (Moyer & Landauer, 1967) is evident in the increasing time required to compare two numerosities (or in the decreasing accuracy of such a comparison) as their difference decreases. In other words, it is more difficult to discriminate a display of 5 dots from that of 6 than from that of 10. When the numerical distance between the numerosities is held constant, the comparison performance decreases as the magnitudes of the numbers increase

(e.g. one compares 1 against 2 faster than 8 against 9) This is usually called the *number size effect* (Parkman, 1971; Moyer & Landauer, 1973). Furthermore, the magnitudes of the compared numbers interact at the semantic level with the question asked to the experiment participants. Banks, Fujii and Kayra-Stuart (1976) found that, for symbolic numbers, subjects who are instructed to ‘choose smaller’ number respond faster when comparing relatively small numbers (e.g. 2 versus 3), and conversely, the subjects are faster to ‘choose larger’ number when the numbers being compared are large (e.g. 7 and 8). This phenomenon is called the *semantic congruity effect*. Finally, the *SNARC effect* (Dehaene et al., 1993) found, among others, in the comparison task, is related with the number-space congruency, and refers to the fact that the answer ‘larger’ is given faster with a right-sided response (e.g. with a press of a button located to the right) than with the left-sided one. As the latter phenomenon is given a prominent attention in present work (in the context of the research question 4), it is discussed in detail in section 2.2.3.

Despite the decades of research, the exact nature of the mechanisms on which the human ability to process number magnitudes is founded is still the subject of debate (see Noël, Rousselle & Mussolin, 2005, for review). The most fundamental, and as of today still unresolved, question is whether humans are equipped with a dedicated cortical machinery that enables them to process the information of strictly numerical nature already at birth. Gallistel and Gelman, Wynn and Dehaene are among the prominent researchers in the field that advocate such a view (Gallistel & Gelman, 1992; Wynn, 1998; Dehaene, Dehaene-Lambertz & Cohen, 1998). There are however scientists who point out that the behaviour observed in human infants that is taken as the evidence of their innate numeracy may also result from the general cognitive capacities that are not specifically numerical (Simon, 1997, 1999; Uller, Carey, Huntley-Fenner & Klatt, 1999; Clearfield & Mix, 2001; Mix, Huttenlocher & Levine, 2002). The perceptual systems that are most often put forward as potentially providing the cognitive basis for the processing of numerical magnitude are reviewed below.

Feigenson, Dehaene and Spelke (2004) describe two ‘core systems of number’ on which our numerical abilities could be based and for the existence of which there is evidence already in infancy. What they refer to as the *core system 1* deals with the approximate representation of magnitude and is engaged in the imprecise processing of large numbers (> 4). This system has two signature properties: the *Weber’s law*, which states that the similarity of the representations of two quantities is the function of their ratio; and the *scalar variability*, which states that the representations become less and less precise with increasing magnitude. These two properties are required for the system to be able to account for the numerical distance and the number size effects. The second system put forward by Feigenson et al., the *core system 2*, is responsible for keeping track of small numbers of individual objects (≤ 4), and, in contrast to the first system, is precise. The crucial difference between these two systems is that in the case of the core system 2, the discrimination performance no longer depends on the ratio between the numerosities, but on their absolute magnitude. The employment of such two systems in operations on non-verbal magnitudes is consistent with the observation that in adults there is a qualitative difference in enumeration performance for numerosities within the range 1–4 and those greater than 4. For the former, the speed and accuracy of enumeration is very high, while for the latter the latencies and error rates climb sharply (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). The ability to recognise the numerosity of small sets immediately and accurately suggests a dedicated enumeration process, which in the literature is referred to as *subitising* (Kaufman, Lord, Reese & Volkman, 1949). It has to be acknowledged however, that there are prominent researchers who question its existence (Gallistel & Gelman, 2000; Cordes, Gelman, Gallistel & Whalen, 2001).

In addition to the two core systems of number described above, Le Corre and Carey (2007) propose a third preverbal representation in which human numerical skills could be rooted, the *set based quantification system*. This system is closely linked to language — more specifically, it is argued to provide the grounding for

the quantifiers and the singular/plural distinction — and ‘explicitly distinguishes the atoms, or individuals, in a domain of discourse from all the sets that can be comprised of them’ (Le Corre & Carey, 2007, p. 398). This system is hypothesised to be available to both non-linguistic primates and preverbal infants.

Finally, there are researchers who argue that ‘numerical’ abilities (at least those of infants) can be explained without resorting to systems which are inherently numerical in nature, as is the analogue magnitude representation system (core system 1 in the nomenclature used by Feigenson et al., 2004). Instead, they postulate that the infant ‘quantification’ skills are based on non-numerical perceptual cues that naturally co-vary with number, such as volume, area, or length for visual stimuli, and frequency, total duration, or acoustic power for auditory stimuli (Clearfield & Mix, 2001; Mix et al., 2002). This argument is plausible in light of the fact that it is not easy to design stimuli in which some continuous perceptual variables would not change together with the numerosity of the set (Mix et al., 2002). This observation has important implications for the possibility of misinterpretation of the results in studies that employ such stimuli. For example, it has been demonstrated that careful control of the perceptual variables causes the children to fail on tasks like comparison of 1 against 2, which should be easy for them, should they possess the abilities of purely numerical nature as it is often presumed (Clearfield & Mix, 1999; Lipton & Spelke, 2004). Furthermore, the continuous nature of the non-numerical perceptual variables in question provides a natural explanation for the observed number size and numerical distance effects. In 2005, Noël et al. admitted that at that time there still was ‘a lack of strong empirical evidences to rule out the perceptual account [for the infant quantification skills]’ (Noël et al., 2005, p. 184).

Irrespective of the actual nature of the processes involved, it is quite clear that humans are able to process numerosity before receiving any kind of formal education in mathematics, and even before they start to learn to count (a detailed discussion of the developmental trajectory is presented in section 2.3.1). In the light of this, learning to count can be viewed as a process of grounding the meaning of

the symbolic codes for numbers (number words, followed by e.g. Arabic notation) in the more primitive, non-verbal representations (Le Corre & Carey, 2007). There is a considerable amount of evidence that such a mapping indeed eventually takes place (Dehaene, 1997; Cordes & Gelman, 2005). Notably, the processing of numbers conveyed in the symbolic form is characterised by the same effects that are found with non-symbolic stimuli, such as the numerical distance and number size effects (see for instance Moyer & Landauer, 1967; Duncan & McFarland, 1980; Schwarz & Stein, 1998; Whalen, Gallistel & Gelman, 1999; Verguts & van Opstal, 2005; Izard & Dehaene, 2008). This suggests that the same underlying representation is used in both cases. Therefore, the study of the ability to count is of paramount importance for the investigation of the human numerical abilities.

2.1.2 Counting

Counting is a process in which the cardinality of a set is determined by shifting the attention from one item of the set to another in some order, establishing at the same time a bijection between the items of the set and the sequence of number words at one's disposal, and retrieving the final result as the number word assigned to the last item (Beckwith & Restle, 1966). In contrast to what can be achieved using the non-verbal quantification abilities that are at humans' disposal (cf. section 2.1.1), the result of a correctly executed counting procedure is always exact. This comes however at the cost of the response time increasing proportionally to the size of the set. As implied by the definition of the counting process given above, correct counting requires having a stable count list, an ability to navigate the set in such a way that all its items are visited exactly once, and, finally, being able to coordinate the articulation of the former with the latter. Since counting is considered one of the fundamental numerical skills and is believed to play a key role in the acquisition of the number concept, it is not surprising that it has been drawing the attention of psychologists for a long time (Piaget, 1952).

Several attempts to decompose the process of counting have been made in order

to aid its study from the point of view of cognition. Already quoted Beckwith and Restle (1966) distinguished three elements of counting: ‘chanting the numerals’, ‘shifting the indicator response’, and ‘grouping objects into those already counted and those still ahead’ (Beckwith & Restle, 1966, p. 437). Somewhat later, Schaeffer, Eggleston and Scott (1974) proposed that the development of the human number knowledge can be seen as a ‘hierarchic integration of six skills’, the first two of which were ‘the counting procedure [...] defined as the consistent coordination of ordered number names and counted objects’ and ‘the cardinality rule [which] states that the last number named during counting denotes the number of objects in an array’ (Schaeffer et al., 1974, p. 358). A more detailed decomposition, which over the years proved itself useful in the study of the development of the counting skill in children and is widely accepted as the classical standard by the researchers in the field, has been proposed by Gelman and Gallistel (1978). Based on the observation that the number words could be viewed just as arbitrary tags, the five *counting principles* of Gelman and Gallistel are formulated as follows:

The one-one principle

each object in the counted set must be tagged with a unique tag;

The stable ordering principle

the tags must be drawn from a stably-ordered list;

The cardinality principle

the last tag used indicates the cardinality of the counted set;

The order-irrelevance principle

the order in which the items are tagged is irrelevant;

The item-kind irrelevance principle

any items can be collected together for counting.

Looking at the counting skill from such a fine-grained perspective allows for a more precise analysis of its development. For example, a child might adhere to the counting principles but using an incorrect, although stable, sequence of number words.

Dismissing such counting as altogether incorrect could obscure the real picture of the child's competence. This illustrates that rather than looking at the acquisition of counting holistically, more insight may be gained by attempting to observe the onset of the child's understanding of the particular counting principles.

An integral part of counting is what Beckwith and Restle (1966) call 'shifting the indicator response', and what during the years when the children learn to count takes the form of touching or pointing gestures. Such gestures are spontaneous, present across the cultures and it is well documented that they improve the counting accuracy throughout the period when the children acquire the skill (see Graham, 1999, for review). Since the role of the pointing gestures in learning to count is another important theme in the present study (cf. research questions 2 and 3), it is discussed in detail in a devoted section (2.2.2).

Assessing the children's counting competence is not easy, as the experimental design and the task instructions may affect the obtained results. Furthermore, the children's behaviour may be ambiguous and different researchers may interpret it differently (Cordes & Gelman, 2005; Le Corre, Van de Walle, Brannon & Carey, 2006). One of the most commonly used tasks in the behavioural study of counting is the *How Many?* task (often abbreviated HM, Gelman & Tucker, 1975). In this task, the experimenter confronts the child with a set of objects and asks a question like 'How many elephants?'. Multiple parameters of the task may be varied, such as the nature of the objects (e.g. actual objects versus their pictures, static versus temporal and visual versus auditory stimuli), homogeneity of the set, or the specificity of the instructions which may explicitly encourage counting or not. A particular variant of this task is the *What's on This Card?* task (WOC), introduced by Gelman (1993) in order to reduce the performance demands and minimise the possibility that the children may misunderstand the task instructions due to their language limitations (Cordes & Gelman, 2005). This task uses cards with stickers and always starts with the presentation of a card containing one sticker. The question asked by the experimenter ('What's on this card?') prompts the child to name the object on the

card (e.g. ‘a bee’). This is followed by the experimenter saying ‘That’s right, it’s *one* bee’ in an attempt to help the child understand that the task is to report the number rather than to label.

Since the children’s accuracy in a task can be affected by its performance demands, it may be argued that the failures in applying the counting principles do not prove the lack of their understanding. In order to alleviate this problem, instead of asking the children to count, one can ask them to watch and assess the correctness of counting performed by someone else, usually a puppet controlled by the experimenter (Gelman & Meck, 1983; Gelman, Meck & Merkin, 1986). A big advantage of this approach is the fine degree to which the children’s sensitivity to the violation of the particular counting principles can be tested by an appropriate orchestration of the puppet’s counting. Possible variations include here the controlled commitment of various types of counting errors as well as employing the counting strategies which are unconventional but correct, like starting to count in the middle of a row.

Quite early on it has been recognised however that being able to execute the counting procedure correctly does not necessarily imply that one understands fully its meaning in the sense of being able to employ counting as a tool to solve a task for which it is required (Schaeffer et al., 1974). In order to study this issue in more detail, additional experimental paradigms have been developed. In the *Point-to-X* task (P2X, Wynn, 1992b), two cards with different numbers of stickers are shown to the children. They are then asked to indicate the card containing the given number of objects (e.g. ‘Can you show me the four balloons?’). Pairing a card containing one item with a card with N items allows one to test whether the child understands that the number word for N refers to a numerosity. In turn, a card with N objects may be paired with the one with $N \pm 1$ objects in order to assess if the child understands which numerosity the word refers to exactly. Another task, *Give a Number* (GN, Wynn, 1990) consists of asking the child to give the experimenter a specific number of items from a large pile at their disposal. It has proved itself to be among the hardest counting tasks, and the fact that children who succeed in the simpler HM

task may very well fail in the GN task is well documented (see for instance Cordes & Gelman, 2005; Le Corre et al., 2006).

At this point it should be evident that the assessment of the counting skills of an individual is not trivial. Many experimental paradigms exist, and up to now there is no consensus among the respected researchers in the field regarding the adequacy of the specific tasks to address particular scientific questions (Cordes & Gelman, 2005). Finding the definitive answers to the outstanding issues will therefore require further work on the part of the experimental psychologists in terms of new robust experiment designs and consistent data analysis procedures. The evidence available at the time of writing which I review in this chapter indicates that the more plausible hypothesis is that the knowledge of the principles governing counting is not innate (in contrast to what has been advertised for many years by Gelman and affiliated researchers) but that the understanding of these principles is acquired through learning to count and thus emerges from the interactions between the structure of the symbolic number representation (the counting list) and the properties of the non-verbal numerical capabilities (which are either acquired earlier or innate) as a result of a link being established between the two (Le Corre & Carey, 2007).

2.2 Embodied Facets of Human Number Knowledge

Having provided the elementary background information about the two aspects of the human numerical knowledge that the present study focuses on — non-verbal processing of the number magnitude and counting — I now discuss the selected lines of evidence that human mathematical thinking is embodied, in other words that it has its roots in our everyday interactions with the world around us. At first, the argument that the most abstract concepts that human mind has conceived — the mathematics — is tightly connected with down-to-earth ‘material’ activities, may seem counterintuitive. This is a consequence of the still widely prevailing Platonic

view of mathematics, according to which this domain of science is transcendent and conveys ultimate truths about the universe. The growing body of research on the cognitive science of mathematical thinking suggests however that this view needs to be revised. It seems that rather than assuming, like Maximillian Cohen, the protagonist of the memorable Darren Aronofsky's 1998 movie *Pi*, that 'mathematics is the language of nature', it may be more appropriate to refer to it as the 'language in which we [...] try to read it' (Dehaene, 1997, p. 252).

From the point of view of this thesis it is important to demonstrate that the human numerical skills have embodied origins, because it is one of the main arguments in support of my attempt to tackle the research questions posed in the introduction with the aid of the developmental cognitive robotics (see chapter 4). In the present section, three examples of the embodied contribution to mathematical thinking are discussed. First, I briefly review the embodied bases for arithmetic proposed by Lakoff and Núñez (2000). Second, I discuss the research which focuses on the ways our hands are involved in counting. Finally, I delve into the evidence that in the human mind the representation of numbers is tightly linked with the representation of space.

2.2.1 The Origins of Mathematical Ideas

The embodied aspects of the human mathematical ideas are the central theme of an influential book written by Lakoff and Núñez (2000). Therein the authors explore how fundamental mathematical concepts like arithmetic, sets, logic, abstract algebra and infinity could be grounded in the every-day sensorimotor experiences.

The approach of Lakoff and Núñez to the analysis of the mathematical ideas from the point of view of embodied cognition is based on three theoretical concepts. The *image schemas* are the basic conceptual primitives that can express ideas (e.g. such as spatial relations) and appear to be universal across cultures. The *conceptual metaphors* reflect the mechanism of understanding abstract concepts in terms of more concrete ones via metaphor. Finally, *conceptual blending* entails forming com-

plex metaphors by putting together simpler ones across the conceptual domains.

In order to show the way in which the properties of arithmetic may come from the sensorimotor experience, Lakoff and Núñez put forward four conceptual metaphors, which they call the *4Gs*. Arguably, these ‘allow human beings [...] to extend arithmetic beyond the small amount that we are born with, while preserving the basic properties of innate arithmetic’ (Lakoff & Núñez, 2000, p. 77). The 4Gs are: *Arithmetic Is Object Collection*, *Arithmetic Is Object Construction*, *The Measuring Stick Metaphor* and *Arithmetic Is Motion Along a Path*. As an example, the first metaphor is briefly described below.

Lakoff and Núñez define the conceptual metaphors by specifying a mapping of certain elements of one domain (the *source* domain) to another (the *target* domain) Arithmetic Is Object Collection metaphor is stated as (Lakoff & Núñez, 2000, p. 55):

- Collections of objects of the same size → Numbers
- The size of the collection → The size of the number
- Bigger → Greater
- Smaller → Less
- The smallest collection → The unit (One)
- Putting collections together → Addition
- Taking a smaller collection from a larger collection → Subtraction

The evidence for this metaphor can be found in the language, for instance a statement ‘Two is smaller than four’ refers to the physical size of numbers, which are not physical objects and thus do not, literally, have a size. More importantly however, Lakoff and Núñez demonstrate the entailments of the Arithmetic Is Object Collection metaphor, showing that the laws governing our interactions with collections of objects correspond exactly to the laws of arithmetic, such as the stability of results,

inverse operations, closure, commutativity, associativity, symmetry, transitivity, etc. (Lakoff & Núñez, 2000, pp. 57–59). For example, for two object collections A and B , either A is bigger than B or B is bigger than A or A and B are of the same size. Similarly, two numbers A and B follow the law of trichotomy. Finally, the theory of Lakoff and Núñez predicts that multiplication and division are cognitively more complex than addition and subtraction, because, in contrast to the latter, defining the former requires referring to both collections and numbers themselves, for example expressing multiplication as performing certain operations on collections a specific number of times. Therefore, Arithmetic Is Object Collection metaphor is an an example of conceptual blending.

Each of the three remaining 4Gs facilitates extending the arithmetic concepts with novel elements. For example, Arithmetic is Object Construction enables conceptualisation of simple fractions. The Measuring Stick Metaphor allows forming a concept of rational number. Finally, Arithmetic is Motion Along a Path provides natural extension for concepts of zero and negative numbers. Lakoff and Núñez argue that taken together, these four metaphors can explain all the most important laws of arithmetic, based solely on the ordinary interactions of humans with their environment.

Although Lakoff and Núñez delve much deeper into the investigation of the embodied aspects of mathematics than presenting the 4Gs, a thorough review of their book would unfortunately be much beyond the scope of this work. Already at this point it should be evident that despite the abstract appeal the mathematical concepts have, they may in fact be tightly grounded in very ‘material’ sensorimotor knowledge.

2.2.2 The Involvement of Hands in Counting

It is well established that pointing to, touching, or moving items during counting is an integral part of the development of children’s number knowledge (Graham, 1999). Children use such gestures spontaneously and several independent studies

have confirmed that this facilitates the counting accuracy (Schaeffer et al., 1974; Gelman, 1980; Saxe & Kaplan, 1981; Gelman & Meck, 1983; Fuson, 1988; Graham, 1999; Alibali & DiRusso, 1999; Carlson, Avraamides, Cary & Strasberg, 2007). According to Schaeffer et al. (1974), preventing the children from pointing severely disrupts their counting: in such case a child most often either emits an indefinite stream of numbers or does not count at all. The studies by Gelman (1980) and Gelman and Meck (1983) provide further evidence for the importance of the actual physical contact with the counted items. The children in these studies experienced more difficulties after the objects being counted have been put behind a transparent cover, which allowed them to point to the objects but not to touch them. Alibali and DiRusso (1999) in turn addressed the issue of active and passive gestures. The former refers to the pointing performed by the child itself and the latter to the pointing performed by somebody else, for example an experimenter-controlled puppet. As it turns out, both active and passive gestures significantly improve the children's counting accuracy over the situation in which the gesture is prevented. The counting competence in children develops over a significant period of time (see section 2.3.2), and accordingly, the contribution of the counting gestures has a clearly developmental character. This conclusion is supported for example by the results of Saxe and Kaplan (1981) who showed that 4-year-old children significantly benefit from pointing, in contrast to 2- and 6-years-olds.

Following the ample experimental evidence for the supportive role of the counting gestures in learning to count, a number of hypotheses concerning the specific nature of this contribution have been put forward in the literature. Among these, three main themes can be distinguished.

The first group of hypotheses views gesturing as a way to *overcome the limitations in the available cognitive resources*. For instance, it can be argued that the gestures may 'externalise' some of the contents of the working memory. According to one of the earliest proposals, pointing while counting is a way to keep track of the counted items (Schaeffer et al., 1974). In order to adhere to the one-one principle (Gelman &

Gallistel, 1978, see section 2.1.2), one has to separate the objects that have already been counted from those that still remain to be counted, otherwise some items may be counted more than once while others may be omitted. A hand pointing toward an object can fulfil the role of an ‘external memory register’, identifying the current object being counted, and, indirectly, all the objects counted so far. This is especially plausible when the counted set is arranged in a way that can be followed by a smooth hand trajectory (such as a single row) and is consistent with the observation that the children find certain arrangements of items easier to count than others (Beckwith & Restle, 1966). The study by Alibali and DiRusso (1999) provides however evidence that keeping track is not the only function of the counting gestures. Should that be the case, the children would count most accurately in the passive gesture condition, that is when they follow a flawless pointing performed by somebody else. The results of Alibali and DiRusso did not confirm this prediction — they found no statistical difference between the active and passive gestures in terms of counting accuracy.

The second category of the hypotheses focuses on the possible *coordinative role* the counting gestures may play in synchronising the production of the number words and matching them with each counted item so that the one-one counting principle is preserved. As pointed out by Fuson (1988), counting gestures combine in a natural way two correspondences: a correspondence in space (between the gesture and the objects), and a correspondence in time (between the gesture and the recited number words). This enables the gesture to perform the role of a ‘cognitive hub’ between the recitation of the number words (characterised only by the temporal aspect) and the objects being counted (characterised only by the spatial aspect). The evidence in favour of this proposal was found by Alibali and DiRusso (1999) in the patterns of the counting errors made by the children in the active and passive gesture conditions. The children in the study committed less coordination errors when they gestured themselves than when the pointing was done by somebody else, suggesting that the active gestures help the children to coordinate the reciting of the number words with assigning them to the objects being counted. A related issue is the observation

that the rhythmical nature of the counting gestures may help the children to better control the shifts of attention and facilitate the correct segmentation of the counting task. In turn, this would help to treat the counted items as separate entities and therefore highlight more prominently the correspondence between the items and the number words. This proposal is consistent with the fact that touching the objects is more effective than merely pointing to them (interestingly, also in the passive gesture condition), as the former is a less ambiguous indication (see Alibali & DiRusso, 1999, p. 52).

Finally, the *social aspect* of gestures should not be neglected. Pointing in particular is an example of a gesture that plays important communicative roles very early in the development (Behne, Liskowski, Carpenter & Tomasello, 2012). Several studies could be named that looked at the gestures from the perspective of social learning in various contexts (see Graham, 1999, pp. 352–353). For instance, the lack of correspondence between the children’s gestures and speech can be taken as an indication that their understanding of a matter undergoes change and therefore that they are ready to learn and would benefit from additional instructions (Breckinridge Church & Goldin-Meadow, 1986; Perry, Breckinridge Church & Goldin-Meadow, 1988; Goldin-Meadow, Nusbaum, Garber & Breckinridge Church, 1993). Such gesture-speech mismatches can be directly observed by the tutor, providing the latter with the feedback about the child’s learning progress (Goldin-Meadow, Wein & Chang, 1992; Goldin-Meadow, Alibali & Breckinridge Church, 1993). The plausibility of this proposal is reinforced by the studies that show that the adults indeed understand the information conveyed by the children’s gestures and that they subsequently use them to adapt the provided instructions (Goldin-Meadow et al., 1992; Perry, Woolley & Ifcher, 1995).

While there is no doubt that the counting gestures allow the children to improve their counting accuracy, as of today the exact mechanism of this contribution has not yet been pinned down, as demonstrated by the variety of the hypotheses presented above. Importantly, two of the four research questions posed in section 1.2 are

directly related to these considerations. By answering the research question 2, I aim to contribute toward the better understanding of the nature of the phenomenon in general. Research question 3 is in turn linked directly with the second group of the hypotheses presented above, as it focuses on the observations about the spatial correspondence between the gestures and the counted items.

Another motor activity that is omnipresent during the development of the mathematical skills in humans across virtually all geographic areas and cultures is finger counting (Fuson, 1988; Dehaene, 1997; Butterworth, 2000). Children around 4 or 5 years of age commonly use strategies involving fingers (such as counting of raised fingers or recognition of the hand configuration) in solving simple arithmetic tasks (Bisanz, Sherman, Rasmussen & Ho, 2005; Siegler & Shrager, 1984). In adults, there is evidence for the involvement of the cortical circuits responsible for finger motor control in numerical processing (see Andres, Seron & Olivier, 2007). Brain imaging studies report an overlap between the areas involved in number processing tasks and in finger movements (Kaufmann et al., 2008), and numerical deficits are often observed together with finger agnosia (Roux, Boetto, Sacko, Chollet & Trémoulet, 2003). These and related findings suggest that finger counting is a prominent example of grounded and embodied cognition, where the construction of abstract internal representations is tightly related to motor actions (Fischer & Brugger, 2011).

2.2.3 The Association of Numbers and Space

Considering the very abstract character of the number concept and the very material aspect of space, it may seem unlikely that these two domains could be associated in the human brain. Nevertheless, the reports providing the evidence of the existence of such a connection can be traced back to as early as the works of the British polymath Francis Galton. In his works published in the 1880's (Galton, 1880a, 1880b, 1881), he describes 'persons who invariably think of numerals in visual imagery', that is who perceive numbers as arranged along spatial layouts. Galton reported that this peculiarity 'originates at an early age' (Galton, 1881, p. 92) and 'consists

in the sudden and automatic appearance of a vivid and invariable “Form” in the mental field of view, whenever a numeral is thought of, and in which each numeral has its own definite place’ (Galton, 1881, p. 88). Although the early reports of this phenomenon were based solely on introspection, the numerous inter-subject consistencies supported their authenticity (Seron, Pesenti, Noël & Deloche, 1992). The phenomenon is now recognised to be the *number-space synaesthesia* (see Kadosh & Gertner, 2011, for a review) and its study is believed to provide useful insight into the nature of the representation of numbers by the brain (Seron et al., 1992). Although number-space synaesthesia affects only a fraction of the human population (Kadosh and Gertner provide the estimate of 20%), it is not the only finding that suggests that we link numbers and space.

The additional evidence for the interactions between numbers and space comes from the timed experiments with simple numerical tasks. Dehaene, Dupoux and Mehler (1990) used the *magnitude comparison task* in order to investigate whether the human subjects compare multi-digit numbers digit-by-digit or by first converting the number to a magnitude representation. In their study, the participants compared the numbers shown on a screen against the fixed standard of 55. To indicate if the number was smaller or larger, the subjects pressed one of the two buttons, either with their left or right hand. The response was to be given as fast as possible but also with as few errors as possible. One of the findings reported by Dehaene et al. is that the subjects who gave the ‘larger’ response with their right hand and ‘smaller’ with their left hand were significantly faster than subjects who used the reversed response mapping (‘larger’ with the left hand, ‘smaller’ with right).

This discovery was further explored by Dehaene, Bossini and Giraux (1993). Here the *parity judgement task* was used, where the subjects have to press one of the two buttons with their left or right hand, depending on the number parity. Nine experiments were conducted in total, looking at the relations between the parity and the magnitude information. The intriguing results from the previous paper were reproduced. Despite the fact that the magnitude information is not relevant in

the parity task, a robust interaction between the side of the response and the number magnitude was found: left-handed responses of the subjects were faster than their right-handed responses for small numbers, and the converse was true for large numbers. Dehaene et al. called the phenomenon the SNARC effect (what stands for Spatial-Numerical Association of Response Codes). Since the publication of the seminal paper by Dehaene et al., the effect has been reproduced by several independent research labs (see Wood, Willmes, Nuerk & Fischer, 2008) and is believed to be a manifestation of an intimate connection between the numerical and spatial domains in the brain (for extensive reviews see Fias & Fischer, 2005; Dehaene & Brannon, 2011). In the year marking the 20th anniversary of its discovery, the SNARC effect is central to the modelling experiments presented in chapters 7 and 8 of this thesis, and therefore it is more than appropriate to review the most important facts about this phenomenon.

Although the SNARC effect is still not understood completely, through the years of research several interesting characteristics of this effect have been established. For instance, it is evident that the SNARC effect depends on the task context. The numbers 4 and 5 can be classified as ‘large’, and thus yield a faster right hand response, if the range of numbers used in the task is 0–5. The same numbers will be considered ‘small’, resulting in a faster response with the left hand, if the range used in the task is 4–9 (Dehaene et al., 1993; Fias, Brysbaert, Geypens & d’Ydewalle, 1996). The flexibility of the SNARC effect extends also to the orientation of the number-space mapping. Among the subjects originating from the ‘western’ cultures, the most commonly found variant of the effect is the one in which the small numbers are associated with the left side of space and the large numbers with the right side of space (Dehaene et al., 1993; Zebian, 2005; Shaki, Fischer & Petrusic, 2009). In other cultures however, the reverse effect — in which the small numbers are associated with the *right* side — can be found (Zebian, 2005; Shaki et al., 2009). Furthermore, it is possible to manipulate the participant’s momentary numbers-space mapping. One way to achieve this is via explicit task instructions — for instance, asking the subject

to imagine an analogue clock face while performing the task leads to obtaining an adequately modified effect, with faster right-hand responses for the numbers smaller than 6, which are located on the right side of the clock face (Bächtold, Baumüller & Brugger, 1998). In a similar vein, Shaki and Gevers (2011) were able to obtain the opposite number-space association by asking the subjects to differently interpret the same set of symbols, which in their language (Hebrew) can function both as numbers (used left-to-right) and letters (read right-to-left). The SNARC effect is susceptible to even more subtle manipulations. Notebaert, Gevers, Verguts and Fias (2006) demonstrated a temporary reversal of the effect by the means of the inducer task paradigm. Shaki and Fischer (2008) showed that in bilingual subjects familiar with both left-to-right and right-to-left reading directions, reading a paragraph in either language affects the subsequent SNARC effect. In a later study by Fischer, Shaki and Cruise (2009), for the same type of subjects a similar modulation of the effect was obtained without prior priming, solely in response to a number word being presented in one of the two languages. In yet another experiment, Fischer, Mills and Shaki (2010) affected the number-space association of the subjects by varying the bias of the horizontal placement of numbers in a text. Finally, in addition to the association along the horizontal axis, the vertical variant of the effect can be obtained as well (Schwarz & Keus, 2004; Ito & Hatta, 2004; Gevers, Lammertyn, Notebaert, Verguts & Fias, 2006; Hung, Hung, Tzeng & Wu, 2008), in which, usually, small numbers are associated with bottom, and the large numbers with top.

The SNARC effect can be evoked not only by one-digit Arabic numbers, which are the most commonly used stimuli in its context. Studies can be found which report the effect for multi-digit numbers (Dehaene et al., 1990; Brysbaert, 1995) as well as the number words (Fias, 2001; Nuerk, Iversen & Willmes, 2004; Fischer et al., 2009). Despite the failure to obtain the SNARC effect for letters by Dehaene et al. (1993), later research showed that SNARC-like effects can be obtained also for non-numerical stimuli. This includes not only the series which are inherently ordinal, like letters, days of the week, months (Gevers, Reynvoet & Fias, 2003, 2004), or

extensively trained sequences of arbitrary symbols (van Opstal, Fias, Peigneux & Verguts, 2009), but also some not obviously ordinal ones, for instance certain musical features (Rusconi, Kwan, Giordano, Umiltà & Butterworth, 2006; Lidji, Kolinsky, Lochy & Morais, 2007), face expressions (Holmes & Lourenco, 2011) and fruits and vegetables (van Dijck & Fias, 2011).

The SNARC effect is not affected by the subject's handedness (Dehaene et al., 1993) and seems to be connected with the relative position of the response, rather than with the response hand. The evidence for the latter comes from the crossed-hands condition in the experiments of Dehaene et al. (1993), the studies which used pointing with one hand as the means to provide the response (Fischer, 2003; Ishihara et al., 2006), and those which avoided using the hands as task effectors altogether (Fischer, Warlop, Hill & Fias, 2004; Schwarz & Keus, 2004).

A slightly different variant of the SNARC effect has been discovered by Fischer, Castel, Dodd and Pratt (2003). Their experiments utilised the *attention cuing paradigm* (Posner, 1980). The task of the subjects was to detect a visual target that appeared either in the left or right side of their visual field. The subjects had to press a single button as soon as they have detected the target, therefore the response was not spatially coded. Fischer et al. found that presenting a number at fixation causes an involuntary shift of attention toward either left or right side of the visual field in the subsequent trial, which was evident in the detection times. More precisely, when a small number was shown before the target, the subjects were faster to detect the target on the left than on the right, and the converse was true for large numbers. Importantly, this happened despite the fact that the numbers did not statistically predict the locations of the targets and that the subjects were made aware of this fact beforehand. The effect is robust and similar to the classical SNARC effect in terms of flexibility — in later research it has been demonstrated that it is possible to evoke the alternative number-space mappings via the task instructions, including reversed and vertical (Galfano, Rusconi & Umiltà, 2006; Ristic, Wright & Kingstone, 2006). In order to distinguish it from the 'canonical' SNARC effect, the

effect discovered by Fischer et al. is sometimes called the *Posner-SNARC effect* (Chen & Verguts, 2010), in reference to the inventor of the attention cuing paradigm.

Another line of evidence for the existence of a link between numbers and space in the brain comes from experiments with patients who suffer from brain damage which results in hemispatial neglect. Such patients, usually having experienced right parietal lesions, exhibit an impairment of attention toward the stimuli occurring on their left side. Zorzi, Priftis and Umiltà (2002) examined such patients using the *physical line bisection task* and the *mental number line bisection task*. In the former, the subjects are asked to point to the middle of a line. In the latter, they are required to estimate the middle of a numerical interval. Zorzi et al. showed that in the physical line bisection task the answers of the patients affected by the hemispatial neglect are biased toward right, consistent with their indifference to the left side of space. In addition however, it was also discovered that a similar pattern is found in the mental number line bisection task — the patients tended to overestimate the middle of the numerical interval, for example responding that the number lying in the middle between 1 and 9 is 8. This suggests that an intimate connection exists between the numerical and spatial representations in the brain. Vuilleumier, Ortigue and Brugger (2004) conducted a study with subjects affected by the same condition using the magnitude comparison task. They found that the participants were consistently slower to compare numbers located left to the reference than to the right, irrespective of the absolute magnitude of the reference (thus the number 6 was responded to significantly faster for reference 5 than for reference 7). Additionally, Vuilleumier et al. considered the clock task used earlier to show the reversal of the SNARC effect in healthy subjects (Bächtold et al., 1998) and found that the performance of the subjects with neglect for large numbers (placed on the left side of the clock face) was particularly affected. It seemed therefore that the hemispatial neglect affects the spatial structure of the mental representation of numbers in the same way as that of space. Recently however, double dissociations between the spatial and numerical tasks considered by Zorzi et al. have been reported (Doricchi,

Guariglia, Gasparini & Tomaiuolo, 2005; van Dijck, Gevers, Lafosse, Doricchi & Fias, 2011). This illustrates that the nature of the interactions between numbers and space is quite complex and the understanding of it is still an ongoing endeavour.

Two important questions about the SNARC effect, and about the spatial-numerical associations in general, that still remain to a large degree open, are when does it appear and what are its origins. The former matter is discussed in the dedicated section 2.3.3 of this chapter, while the current understanding of the latter is briefly reviewed below. Quoting some of the most prominent researchers in the field, although ‘it is clear that the origin of the relationship between numbers and space is multifaceted and cannot be reduced to one single mechanism’ (Fias, van Dijck & Gevers, 2011, p. 145), ‘there is [...] good evidence that the direction of the number line is [to a large degree] culturally determined’ (Fias & Fischer, 2005, p. 52). In fact, the cultural factors were hypothesised to play an important role in shaping the SNARC effect as soon as it has been discovered. In their seminal study, Dehaene et al. (1993) included an experiment aimed at assessing if the subjects’ *reading direction* interacts with their spatial-numerical association (see experiment 7, Dehaene et al., 1993). Although the expected reversal of the SNARC effect in the Iranian subjects as a group was not found, a correlation was observed between the amount of exposure to the ‘western’ culture the subjects have had received (in terms of the number of years since their settlement in France and of the age at which they acquired the second language) and the slope of their SNARC effect, with some of the individuals exhibiting the ‘reversed’ SNARC. It is plausible to expect that the reading direction may be contributing to shaping the person’s number-space association, because, as Göbel et al. put it, ‘acquisition of writing and reading in a particular language entails uncounted hours of spatially directional perceptual and motor routines’ (Göbel et al., 2011, p. 551), and this is of course likely to leave a mark on one’s cognition. As pointed out earlier in the discussion of the properties of the SNARC effect, numerous studies confirmed that the scanning patterns connected with reading correlate with SNARC both in long and short term.

Crucially however, the *correlation* between the reading direction and the orientation of the SNARC effect cannot be taken as the evidence for *causation*. Furthermore, there are other spatial biases of various kinds persistent across the cultures to which young people are exposed. The examples of such biases are the *directionally-oriented aids* omnipresent in the mathematics education, such as rulers, axes of charts and number lines (Fueyo & Bushell, 1998), as well as the *finger counting habits* (Butterworth, 2000). Especially the latter drew a special attention of some of the researchers, since a link between the finger counting strategy employed by an individual and their magnitude of the SNARC effect has been found by Fischer (2008) in Scottish participants. Fischer and Brugger argue that finger counting is ‘a prime candidate for the origin of directional SNAs [spatial-numerical associations] and their cross-cultural variation’ (Fischer & Brugger, 2011, p. 3), pointing out the problems with assuming the reading direction to be the main contributor. First, the reading-induced SNARC is easily manipulated (Fischer et al., 2009, 2010). Second, in some cultures the opposite orientation of the ‘default’ SNARC effect to the the dominant reading direction can be found (Ito & Hatta, 2004), and the associations orthogonal to the reading direction exist as well (Schwarz & Keus, 2004; Gevers, Lammertyn et al., 2006).

The most compelling evidence that one’s direction of the spatial-numerical association cannot be shaped solely by reading is the fact that spatial biases can be found in children before they actually start to learn to read. Tversky, Kugelmass and Winter (1991, p. 519) review research which shows that the children’s perceptual exploration is spatially biased in a culturally-specific way already during the early years of learning to read. Similarly, Opfer and Furlong (2011) have found in their experiments with pre-reading American preschoolers that the children spontaneously counted from left to right, and that the numeric information consistent with the left-to-right ordering (dominant in the culture of the participants) helped the children to encode spatial relations with higher accuracy. When the ordering was opposite, the children ‘experienced tremendous difficulty’ (Opfer & Furlong,

2011, p. 691). The results led the researchers to conclude that ‘numeric effects on spatial encoding develop far too early to be caused by reading practice or schooling’ (Opfer & Furlong, 2011, p. 682). Instead they hypothesised that ‘some of the interesting links widely observed between numeric and spatial coding [...] are laid down in early childhood as children repeatedly engage in the physical action of counting’ (Opfer & Furlong, 2011, p. 691). An identical hypothesis underlies my research question 4, which is an attempt to contribute additional evidence toward the still-open discussion about the ontogeny of spatial-numerical associations.

It is important to acknowledge that, regardless of their actual nature, the cultural factors may not be the sole determinant of the spatial-numerical associations and that alternative hypotheses about their origins also exist (see Fias et al., 2011, for review). Some researchers speculate that numbers and space may become associated at the conceptual level as the result of people talking about the spatial and numerical concepts using the same language (see Srinivasan & Carey, 2010, for similar considerations in the context of space and time domains). Other recent experimental results suggest that the tendency to relate the representations of number to spatial length may even be innate (de Hevia & Spelke, 2010). Furthermore, there is an increasing amount of evidence that the mechanism behind the number-space associations may be altogether different than previously thought, with the serial ordering in the working memory being the crucial factor (van Dijck & Fias, 2011; Fias et al., 2011).

Concluding, the body of evidence for the embodied origins of the mathematical concepts is convincing and constantly growing. The studies that demonstrate the involvement of the hand motor circuits in counting and the existence of a connection between the representations of numbers and space in the brain, together with the theoretical considerations of Lakoff and Núñez (2000) suggest that it should be promising to supplement the efforts in computational modelling of mathematical cognition with an artificial, robotic body.

2.3 The Development of Human Numerical Skills

The study of the development of the mind is important for several reasons. First, gathering the knowledge about the ontogeny of a mental capability is one of the crucial steps toward understanding it as a whole. Second, knowing how the development of a cognitive skill typically progresses allows us to diagnose its dysfunction. In the case of a typical development, understanding how it works holds a promise of designing more efficient teaching guidelines and refining the existing educational curricula. Finally, the findings concerning the development of cognition in humans are of great interest for roboticists who envision creating autonomous, robust and truly intelligent machines, and who believe that the rules governing human learning and development holds the key to finally achieving this long-lasting dream (see section 4.2).

In the following subsections I review the literature which focuses on the developmental trajectories of the three aspects of human numerical skills that have been discussed earlier in this chapter. The preverbal magnitude processing is considered first, followed by counting, and, finally, the spatial-numerical associations. At the end, this short review is summarised by a sketch of a timeline of the development of the early human mathematical knowledge.

2.3.1 The Development of Non-verbal Magnitude Processing

First, I focus on the developmental trajectory of the elementary non-verbal quantification abilities, that were introduced in section 2.1.1. To investigate the development of the pre-verbal numerical skills in humans, specialised experimental paradigms are necessary, because the traces of such skills can be found in children very early on in the development (i.e. hours after birth), when not only any kind of communication with the subject is ruled out, but also the repertoire of the behaviours that can be observed is extremely limited. Most of the existing studies therefore are based on methods like preferential looking, dishabituation or violation of expecta-

tion (Colombo, Brez & Curtindale, 2013). These approaches exploit the fact that infants are sensitive to novelty, and that it is possible to observe if they experience something as novel from various cues in their behaviour, such as the gaze direction or duration.

Human infants can, to a certain extent, discriminate numerosity already at very early age. The first studies of the development of this ability focused on establishing its onset and its limits. Antell and Keating (1983) found that as early as 21 to 44 hours after birth, neonates can discriminate between the sets of two versus three dots. The ability to detect changes in small sets presented visually, as long as the number of items does not exceed about 4, has been confirmed in multiple studies with infants of ages ranging between 3½ to 13 months (Starkey & Cooper, 1980; Strauss & Curtis, 1981; Van Loosbroek & Smitsman, 1990). Later on, it has been demonstrated that, in addition to being able to discriminate between small sets, 6-month-old children also notice a difference between much bigger numerosities, provided that their ratio is sufficiently large. In studies by Xu and Spelke (2000) and Xu, Spelke and Goddard (2005), infants were able to discriminate between large numerosities when their ratio was 1 : 2 (8 versus 16 and 16 versus 32), but not 2 : 3 (8 versus 12 and 16 versus 24).

The ability to discriminate numerosity does not seem to be exclusive for the visual stimuli presented in the form of the static images, suggesting that abstract, amodal representations are involved. In the study by Wynn (1996), 6-month-old children discriminated visual stimuli with a temporal aspect, namely the sequences of jumps of a puppet (2 versus 3 jumps). Bijeljac-Babic, Bertoni and Mehler (1993) monitored the sucking rate of neonates (3 to 4 days old) and have found the evidence of dishabituation using two- and three-syllable pseudo-word utterances as the stimuli. Lipton and Spelke (2003) tested 6- and 9-month-old infants using auditory sequences, and discovered the ratio effect very similar to the one reported by Xu and Spelke (2000) for static images, furthermore showing that the precision of numerosity discrimination increases along with the age. The study was followed

up by the same authors (Lipton & Spelke, 2004) extending the findings to smaller numerosities. Other researchers have also looked at the auditory-visual cross-modal matching using various experimental paradigms (Starkey, Spelke & Gelman, 1983; Kobayashi, Hiraki & Hasegawa, 2005), although their findings turned out to be fragile and difficult to reproduce (Mix, Levine & Huttenlocher, 1997; Moore, Benenson, Reznick, Peterson & Kagan, 1987).

In an influential study Wynn (1992a) demonstrated that at the age between 4 and 5 months children are sensitive to the elementary numerical transformations on small sets, such as addition and subtraction. The experiments used the violation of expectation paradigm with the children's gaze duration acting as the index of the level of surprise. For example, for the addition experiment, the children were shown a puppet which was subsequently occluded by a screen. The children then saw the experimenter's hand placing another puppet behind the screen, after which the screen was lowered revealing one of the two possible outcomes. In the *expected* outcome, two puppets appeared behind the screen ($1 + 1 = 2$). In the *unexpected* outcome, after the screen was dropped, only one puppet remained ($1 + 1 = 1$). The analysis of the gaze duration revealed that the children showed surprise when faced with the unexpected outcomes (e.g. $1 + 1 = 1$ or $2 - 1 = 2$), providing evidence that already around 4 months of age they form arithmetically correct expectations toward the results of simple addition to and subtraction from small sets. The experiment has been replicated by several independent researchers, who showed in addition that the children are not confused by the motion of the objects (Koechlin, Dehaene & Mehler, 1997) and that in the considered set-up they do not attend to the object identity at 5 months of age (Simon, Hespos & Rochat, 1995). The latter means that the children detect the violation of the numerical outcome of the experiment, but not a physically impossible transformation of one object into another (e.g. a duck + a duck = a duck and a car). This persists until they are 10 months old, with 12-month-old children finally demonstrating the sensitivity to the object identity as well (Xu & Carey, 1996).

The decades of experimental research have provided a lot of valuable data on the development of the non-verbal magnitude processing in humans, but, as already mentioned in section 2.1.1, the debate about the exact nature of the mechanisms underlying it is not fully resolved yet. Independently of establishing what preverbal quantification systems are actually at our disposal, several questions about the rules governing the interactions between them remain open. Among the most puzzling facts is the interaction between the set size and the set sizes ratio that is found in the infants' discrimination of the static visual and the temporal auditory stimuli. That is, for a fixed ratio between the numerosities, the children succeed on the large sets but fail on the smaller ones which supposedly should be easier for them to discriminate (Xu & Spelke, 2000; Clearfield & Mix, 1999; Lipton & Spelke, 2004). Providing satisfactory explanations for such phenomena remains the part of the future work not only for the experimental psychologists but also for the theorists and cognitive modellers.

Of course number magnitude comparison can also be studied using the symbolic stimuli (see for instance Sekuler & Mierkiewicz, 1977; Duncan & McFarland, 1980). This alleviates the problems related to the perceptual variables that correlate with number, however the obvious requirement is the prior acquisition of an appropriate repertoire of verbal codes for numbers by the subjects. As mentioned in section 2.1.1, this is believed to be achieved through the process of learning to count, which is outlined in detail in the next section.

2.3.2 The Development of Counting

As discussed in section 2.1.2, a crucial ingredient of counting is the recitation of the number words. The studies of Fuson, Richards and Briars (1982) were entirely devoted to the investigation how children's mastery of the counting list develops with age. They distinguished two overlapping phases of the development. The *initial acquisition phase* typically begins around 2 years of age and lasts more or less until 6–7 years of age. During this period the children learn the conventional

sequence of number words, and begin to use it to count objects. During the later *elaboration phase*, the children refine their understanding of the structure of the counting sequence — how it is divided into the individual number words and what are the relations between them. Fuson et al. argue that when this elaboration phase is completed ‘the words in the sequence themselves become items which are counted’ (Fuson et al., 1982, p. 33), and that this subsequently scaffolds the learning of the rules of arithmetic. Therefore, the main qualitative difference between the two phases is that in the former the ‘sequence functions as a single, connected, serial whole from which interior words cannot be produced independently’, while after the latter the counting list ‘has the structure of an associative chain’ (Fuson et al., 1982, p. 34) in the sense that children can for example recall the counting sequence between two arbitrary numbers and count in the reverse direction. An important observation is that because the learning of the counting list spans over several years, different parts of it can be at different stages of the development at the same time. Furthermore, the early parts of the counting list typically precede in development the later ones (Fuson et al., 1982, p. 34).

Because it is the initial acquisition phase that overlaps with the children’s learning to count, it is appropriate to review it in a bit more detail. According to Fuson et al. (1982), learning the number words up to ‘twenty’ is, in principle, a serial recall task, in other words, this part of the counting list has to be learnt by rote. Above ‘twenty’, the structure of the number words becomes regular, changing the nature of the task. The children learn to distinguish the number words from other words very early and intrusions of other words in the counting context are rare and usually limited to the letters of the alphabet (probably because they are learnt in a similar way to counting words, Fuson et al., 1982, p. 35). During the initial acquisition phase, while the children learn number words up to approximately 30, the counting sequence used by a child can be typically divided into three portions. In the initial, *stable conventional* part, children recall the number words correctly. This is followed by a *stable non-conventional* portion consisting of number words which do

not follow the correct ‘adult’ sequence, but are used consistently by the particular child in repeated trials. Finally, the counting list of the child reaches the *nonstable* portion characterised by no or little inter-trial consistency. The length of the stable conventional portion increases with age. According to the data gathered by Fuson et al., a considerable extension of this portion occurs on average when the children are between 4½ and 5 years old. They report the mean length of the single best trial in the number list recitation task for children between 4 years and 6 months and 4 years and 11 months to be 29.59 (Fuson et al., 1982, table 2.2). This is accompanied however by a large standard deviation (28.19) which is the consequence of a vast between-subjects range (reported to be 12–100). Overall, Fuson et al. report ‘extreme’ variability within age groups, with some 3-year-olds having longer conventional portions of the counting list than some 5-year-olds. This variability persists ‘into the early grades of the elementary school’ (Fuson et al., 1982, p. 39). Based on longitudinal data, in which an interaction between the level of consistency and time has been found, Fuson et al. distinguish two aspects of the stable portion acquisition process: the ‘extension of the sequence and consolidation of this extension so that it is always produced’ (Fuson et al., 1982, p. 42). It takes approximately five months to thus ‘consolidate’ the most recent extension.

The non-stable conventional portion of the counting list produced by children is usually formed by the omission of certain number words from an otherwise correctly ordered sequence, which is consistent with the serial recall hypothesis (Fuson et al., 1982, p. 43). Initially, the children omit multiple number words most frequently in the range from ‘thirteen’ to ‘seventeen’. Over time, the ‘gaps’ are filled-in with correct number words. The English-speaking children tested by Fuson et al. experienced particular difficulties with the number word ‘fifteen’, which often turned out to be the last remaining omission. Fuson et al. hypothesise this may be caused by the irregular structure of this word in comparison to the surrounding words (i.e. ‘fifteen’ is not ‘fiveteen’). Above ‘twenty’ the structure of the stable non-conventional sequence is somewhat different, likely resulting from the children’s struggle with

mastering of the decade system (for details see Fuson et al., 1982, p. 46).

Learning the correct sequence of number words, although important, is of course only one of the several aspects of learning to count. Another distinctive dimension of the counting competence are the counting gestures, discussed in section 2.2.2. As mentioned therein, such gestures appear to be particularly helpful during a specific stage of the development of the counting skill. The investigation of the developmental aspect of the contribution of pointing gestures to the accuracy of counting was the topic of the study by Saxe and Kaplan (1981). They attempted to establish the developmental differences in the children's reliance on gestures in counting by comparing their performance in two conditions: one in which the children were allowed to point, and one in which the experimental design prohibited it. Having tested the children in three different age groups (2-, 4-, and 6-year-olds) Saxe and Kaplan discovered that only for the middle group the pointing significantly improved their counting accuracy. 6-year-olds scored near ceiling in both conditions and actually rarely used gestures when these were allowed. 2-year-olds in turn counted inaccurately regardless of their use of the gestures. This suggests that the supportive role of pointing in learning to count has a clearly developmental character, and, while they are useful during the period of the acquisition of the skill, by approximately 6 years of age the cognitive aid that comes from the gestures is no longer needed.

Somewhat later, Graham (1999) looked at the role of the spontaneous gestures in children's counting in the course of their development as well as at the relationship between the gestures and the one-one counting principle. In her study, 2-, 3-, and 4-year-olds participated in the HM task and in a puppet counting assessment task. Since the focus was on the spontaneous use of pointing, the HM experiment did not include any specific instructions nor feedback regarding the children's gesture. In the puppet task, the subjects were asked to tell whether the puppet's counting was correct, and if not, to explain the reason. The children's explanations were then analysed for their reference to the number words and to the gestures. In

Graham's study, nearly all children pointed as they counted, even though no specific encouragement was given. It was found that with age, less and less mismatches between the children's gestures and speech occur, in other words the children got better and better in coordinating pointing with the recitation of number words as they grew older. In addition, the adherence of the children's pointing and speech to the one-one principle was measured separately. Interestingly, it turned out that the 2- and 3-year-old children adhered to the one-one principle in gestures more often than in speech. Thus, it seems that the one-one principle appears first in the gestures and is then transferred to speech. The analysis of the puppet task results provided additional evidence that the children's sensitivity to the gestures increases with the age, as the children referred to the puppet's gestures in their explanations the more often the older they were. The results of Graham suggest that the motor competence acquired along with the ability to point to the objects according to the rules of counting may facilitate the acquisition of the conceptual understanding of the counting principles.

A large body of the experimental studies looked at the acquisition of the counting skill from a holistic perspective. Already from early on, these studies attempted to establish the approximate age at which the understanding of the particular counting principles emerges. Schaeffer et al. (1974) studied the development of the counting competence in children between 2 and 6 years of age. They focused on three aspects of counting: the cardinality rule, the counting procedure and what they called the ' $x+1 > x$ principle', which meant understanding that, for instance, 6 denotes a larger set than 5. By analysing the age of children who were able to count items correctly but did not apply the cardinality rule (i.e. did not recognise that the last word they used in counting denotes the size of the set), they showed that the understanding of the latter emerges between $3\frac{1}{2}$ and $4\frac{1}{2}$ years of age. In addition, Schaeffer et al. conducted an experiment looking at the importance of the ability to build a spatial plan in order to execute the counting procedure correctly. They used homogeneous and heterogeneous arrays of objects, in the latter case grouped together by kind

(e.g. a set of toy tigers, cars and buttons). They have found that when counting heterogeneous arrays, children (around 4½ years old) autonomously tended to count items accordingly to the sub-groups (i.e. first all cars, then all tigers, etc.), what apparently facilitated their ability to build a correct motor plan for counting, as it enabled them to overcome their current limit of counting in the homogeneous case, lifting the average accuracy from 51% to 95%. These results suggest that the motor planning abilities are an important factor affecting the correctness of counting.

Beckwith and Restle (1966) and Potter and Levy (1968) also investigated the impact of the spatial arrangement of the counted objects on the performance of the counter. In the former study, the speed and the accuracy of counting of children aged between 7 and 10 years was assessed on homo- and heterogeneous arrays of objects arranged in four different patterns: a line, a circle, a rectangle and a random one. The results suggested an increasing difficulty of these patterns in the following order: rectangle, line, circle, and random, in terms of both the counting latency and the number of errors made. Schaeffer et al. (1974) later hypothesised that this may be caused by the fact that when the objects are arranged in a rectangle, it is easier for a child to build a consistent and easy-to-follow spatial plan for counting, thus helping the child to remember which objects have already been counted. In addition, the study by Potter and Levy (1968) revealed an interaction between the spatial arrangement and the homogeneity of the sets in younger children (2 to 4 years old), indicating that the ability to exploit the arrangement of objects and their category in order to build a motor plan for counting increases with age, as does the capacity to deal with the multiple (potentially contrary) cues from these two sources.

Using her own elaborate decomposition of the counting competence, Gelman (1980) showed the emergence of the understanding of the one-one principle between 3 and 5 years of age. The children were the better in applying the principle the older they were, often achieving near-perfect accuracy for numbers below 5 already at the age of 4 ('near-perfect' was defined by Gelman as using $N \pm 1$ tags for a set of N

objects). For numbers up to 5, even 3-year-olds used $N \pm 1$ tags in more than 80% of the trials. Gelman suggested that the primary reason for the children's errors were the performance demands, arguing that the most of the children's mistakes were related to the coordination of the pointing gestures with the production of the number words. Concerning the stable order principle, more than 80% of the 3-year-olds already used a consistent list of tags in all trials (numbers up to 19), the fraction being of course larger for older children. The data on the acquisition of the cardinal principle provided evidence that the adherence to this rule is a function of both the age and the set size — the children of all tested ages adhered to the principle less frequently for larger sets, including the 5-year-olds, who applied the rule 100% of the time for numbers up to 4. Gelman reports that in the study the variations of the object kind yielded 'little, if any' (Gelman, 1980, p. 62) effect on the counting performance of the tested children. One of the conditions which did affect the counting accuracy was whether the children were allowed to touch the items being counted or not. For the 3-year-olds, the counting performance of the same 3-item sequence dropped from 87% to 49% after putting the items behind a transparent cover. In contrast, the performance of 4-year-olds in the same experiment dropped only for the set sizes greater than 7.

Later study by Gelman and Meck (1983) used the error detection paradigm instead of the HM task in an attempt to separate the effects of the performance demands of the task from those of the children's conceptual competence on their counting accuracy. Here the children assessed the correctness of the counting of a puppet which was controlled by the experimenter so that the child's understanding of the particular counting principles could be investigated. Gelman and Meck looked in particular at the one-one, stable order and cardinality principles, discovering that in some cases even 3-year-old children are able to detect the violation of these rules, although they are significantly worse than the 4-year-olds in the case of the first two rules. This provides evidence that the errors children make in more difficult counting tasks may result from the task performance demands rather than from the child's

lack of the understanding of the counting principles. An extended version of the study published by Gelman et al. (1986) confirmed these findings and in addition indicated that the social factors, such as the unambiguity of the task instructions, influence the child's performance, the stronger the younger the child is. The effect of the type of the question asked on the child's counting performance was found in later studies as well (Cowan, Dowker, Christakis & Bailey, 1996).

In order to investigate the onset of the children's understanding of the cardinality principle, Wynn (1990) conducted a series of experiments in several paradigms, including the standard HM task as well as the GN task. The central issue of the study was to address the ambiguity of the behaviours which in the earlier works were taken as an evidence for the children's understanding of the cardinality principle. For example, Wynn argued that putting the emphasis on the last number word may simply result from a meaningless imitation of the adults or just be an indication of the end of the procedure. Such doubts, reinforced by the experimental results showing the children's failures in tasks requiring counting at the relatively late age (see for instance Schaeffer et al., 1974) called for a more scrutinised investigation of the issue. Combined results of the performance of children in the HM and GN tasks in the Wynn's study indicated that the true understanding of the cardinal principle can be observed around 3½ years of age. For example, participants able to reliably succeed in the GN task appeared only in the oldest considered age group (3 years 6 months on average). Another important finding was that the patterns of children's success in the GN task showed a gradual progression through being able to give a subsets of up to 1, 2, and 3 items. The children who could construct a set of 4 items, succeeded for higher numbers as well. Crucially, the children's limit in the GN task was not connected with their ability to count, as all children could execute the counting procedure correctly up to numbers higher than their GN task limit. The highest number for which the children succeeded also correlated significantly with their age. This led Wynn to conclude that the 'children learn the meanings of smaller number words before the larger ones, even for number words well within

their counting range' (Wynn, 1990, p. 180), in other words, that the knowledge of numbers seems to be separate from the knowledge of counting.

With the aim to study more closely how the children's understanding of the meanings of the small number words develops, Wynn followed her results up in a longitudinal study (Wynn, 1992b). One of its aims was to establish how long does it take a child to go through the individual stages of understanding the early number words. The children's competences were assessed using the HM, GN and P2X tasks. The results from the latter indicated that already early on in the numerical development (i.e. when they reliably pass the GN task for 1-item sets only) the children do know that the larger counting words refer to numerosities, although they do not know to which numerosities they refer to exactly. Moreover, the results from the GN and P2X tasks were consistent in that the limit of the children's performance in giving the sets of items predicted their ability to distinguish between the sets with consecutive numbers of items. The experiment results showed also that a significant amount of time (of the order of one year) is needed for a child to advance from knowing only the meaning of 'one' to fully mastering the cardinal word principle, indicated by the success in the GN task for numbers 4 and larger. Furthermore, the consecutive phases were confirmed to be distinctive and prolonged, with the average duration of the first two phases ('knowing' 1 and 2) reaching 5 months. Importantly, the analysis of the age of the children at which they progress to the next 'number stage' revealed large individual variations. The mean age at the transition from a group to the next one did not correlate significantly with the group number. Furthermore, according to the Wynn's sample, at the age of just above 3 years the children could belong to any of the four distinguished competence groups. This suggests that the absolute age of a child might not be the best index of their number knowledge.

It may be argued that the tasks used by Wynn (1992b), especially the GN task, are excessively difficult for young children (Cordes & Gelman, 2005). Therefore, in parallel, other researchers focused on formulating the experimental paradigms that

would further reduce the posed performance demands. Gelman (1993) tested the children between 2½ and 3½ years of age on the WOC task with set sizes ranging from 1 to 7. In the study, the two oldest groups of children (i.e. 3- to 3½-year-olds) succeeded in at least 80% of the trials for numbers up to 6. Moreover, the scores in the WOC task were consistently higher than those obtained in the HM task across the children’s age group and the set sizes, supporting the claim that the former task is more sensitive to the children’s early numerical competence.

Le Corre, Van de Walle, Brannon and Carey (2006) attempted to reconcile the apparent conflict between the results of Wynn (1992b) and those of other studies (for instance Gelman, 1993; Gelman et al., 1986). The Wynn’s study suggested that the children’s numerical understanding undergoes a significant qualitative change which is not complete until around 3½ years of age. The other studies, which used arguably less-demanding tasks (Cordes & Gelman, 2005), provided the evidence for the children’s knowledge of the counting principles at earlier ages than predicted by Wynn. Le Corre et al. pointed out that this discrepancy may result from a hidden pitfall connected with the standard practice of grouping the children by age. According to the results of Wynn, the age at which the children acquire the understanding of the small numbers, in the sense of the success in the GN task, is highly variable. Therefore, grouping them by age carries a high risk of mixing in one group the children with qualitatively different numerical competences. Le Corre et al. argued that it may be more appropriate to determine the current developmental stage of a child using the GN task rather than their age, and then analyse the children’s performance in the other tasks as the function of the former.

Le Corre et al. (2006) introduced a term *N-knower* to indicate the biggest number for which a child reliably succeeds in the GN task (e.g. ‘2-knower’). They used the term *CP-knower* to refer to a child who succeeds on the GN task for arbitrary large numbers (CP stands here for the ‘cardinality principle’). The term *subset-knower* refers to all children who have not yet become CP-knowers. In their study, having established the children’s *knower levels* using the GN task, they subsequently

analysed their performance on the WOC task and an easy variant of the puppet counting assessment tasks. Although their results provided traces of evidence that the GN task may indeed slightly underestimate the children's numerical competence, it nevertheless proved to be a good predictor of the children's performance on the other tasks, and therefore a good index of the children's numerical knowledge. The individual children's knower levels were consistent across the tasks, with the maximum difference of one level. Crucially however, qualitative differences in performance of CP-knowers in contrast to the subset-knowers have been found. For instance, in the puppet assessment task, only the CP-knowers scored statistically above chance, despite that the employed variant of the task could have been solved simply by using the *last word principle* (i.e. simply matching the last word used with the desired result of counting). Overall, the study provided the evidence against the argument that the poor performance of young children in the studies of Wynn (1990, 1992b) was caused primarily by the task performance demands. Instead, the results supported the view that the subset-knowers indeed possess a qualitatively different understanding of how counting represents number than the CP-knowers, and that the transition from a 0-knower to a CP-knower takes more or less one year.

In later work, Le Corre and Carey (2007) focused on the relationship between the preverbal quantification systems and the acquisition of the meaning of the number words through counting. In particular, they explored the proposal that the principles governing the verbal counting are acquired by children as the result of mapping the latter onto the former. The degree to which the children map the number words onto their preverbal quantification system was assessed using the number estimation and non-verbal ordinal comparison tasks. In the former, the subjects have to provide a guess of the size of the presented set without counting; in the latter, they have to indicate the larger of the two presented sets, also without counting their elements. Both tasks require therefore the use of the non-verbal quantification system (see section 2.1.1). The idea was that the more precise mapping between the preverbal representations and the number words has been established by a participant, the

more accurate their responses in those tasks will be. The performance of the children on the tasks was analysed as a function of their number-knower competence (Le Corre et al., 2006). The first main finding of Le Corre and Carey (2007) was the confirmation of the existence of a distinct, albeit relatively short, 4-knower phase (in the original studies by Wynn all children who succeeded in the GN task for 4, were at the same time CP-knowers). Second, the results of the study provided further evidence that being able to execute the counting routine is not equivalent to acquiring the meaning of all the numerals within the counting range. Although all subset-knowers in the study could count at least up to 10, they could not estimate the sizes of sets larger than 4. The frequency of their use of the number words for large sets did not correlate with the sizes of the sets. This suggests that these children have not yet grounded the meaning of the number words beyond ‘four’. Third, the study revealed that the children’s inability to estimate the sizes of the sets larger than 4 prevails for up to 6 months after they become CP-knowers, that is when they are proficient in constructing sets of objects of arbitrary size. This clearly dissociates the acquisition of the counting principles from mapping the number words onto the non-verbal representations of magnitude. The age at which this mapping finally occurs was estimated to be around 4½ years. Finally, the results of Le Corre and Carey (2007) did not indicate the scalar variability of the children’s estimates for small sets. This suggested that, in non-verbal quantification, the children could not rely solely on the magnitude representation with commonly assumed analogue properties.

2.3.3 The Development of Spatial-Numerical Associations

Despite the great amount of attention the SNARC effect has been receiving from the research community since its initial discovery, there is only a handful of experimental studies which focus on its developmental trajectory (Wood et al., 2008).

Historically, the first study that provided the developmental data regarding the SNARC effect was published by Berch, Foley, Hill and Ryan (1999). The experiment

subjects were American 2-, 3-, 4-, 6- and 8-graders (with mean ages 7:9, 9:2, 9:10, 11:8, 13:7, where $n:k$ means n years and k months), tested on the parity judgement of the Arabic numbers. A significant SNARC effect was found in the children from the third grade on, that is only after two years of formal education. This result is consistent with the hypotheses about the crucial role of the cultural factors, such as the reading direction and the mathematical educational aids, in the shaping of the spatial-numerical associations of an individual (cf. section 2.2.3). Berch et al. report however that the response times of the second-graders were characterised by a large between-trial and between-subject variability, and speculate this may have obscured the SNARC effect in this group (Berch et al., 1999, p. 305). This illustrates a major problem with the experimental investigation of the development of the SNARC effect — the classical experimental paradigms in which the effect is robustly found, depend crucially on the subjects' familiarity with symbols (such as Arabic numbers) and concepts (such as parity) that are acquired at quite late age and through explicit tutoring.

The studies by Bachot, Gevers, Fias and Roeyers (2005) and by Zebian (2005), provide additional evidence that the SNARC effect can be found in children early on into their formal education, although the trajectory of the development is not analysed therein. Bachot et al. were interested whether the spatial representation of numbers is affected in children with visuospatial disability (VSD). They tested the children for the SNARC effect in the number comparison task and contrasted the results with those of a group of healthy children that were matched for age, gender, and verbal IQ scores. The children's ages ranged from 7 to 12 years, with means 9.29 years (VSD group) and 9.24 years (control group). A significant SNARC effect was found for the control children, taken as a group. Unfortunately the data were not analysed as the function of the children's age. In turn, Zebian (2005) performed an extensive cross-cultural study of the SNARC effect and its connection with the dominant reading direction. One of the tested groups consisted of Arabic-English bi-literate children, aged between 8 and 12 years. These children were Arabic

native speakers who received formal education in English as the second language, as well as instruction in maths and science in English. The task used by Zebian was constrained by the requirements of the various groups of participants that were tested (which included illiterates familiar with numerals), and was a modified version of the same-different task (Dehaene & Akhavein, 1995). Only symbolic numbers were used (in the Eastern Arabic format), and the response was given orally ('yes'/'no') instead of bi-manually. A significant reversed SNARC effect was found in the children group, i.e. the responses were faster if the larger number was presented on the left. Furthermore, the magnitude of the effect was larger than in the tested bi-literate adults.

Later, van Galen and Reitsma (2008) demonstrated that the SNARC effect can be found in children already in their seventh year. In a developmental study, they tested children from Dutch first, second, and third grade (on average 7.0, 8.0, and 9.2 years old) as well as adults. Two tasks were employed, the classical magnitude comparison task and the target detection task, in which the number magnitude is not relevant (the Posner-SNARC effect, see Fischer et al., 2003, and section 2.2.3). The standard version of the SNARC effect was found in the magnitude comparison task in all tested groups, including the youngest children. This was the first documented case of the effect at such an early age. The youngest participants however still have been exposed to a certain amount of formal education, as all children in the study 'were tested approximately 3 months before the end of the school year' (van Galen & Reitsma, 2008, p. 103). In contrast, in the target detection task, the biasing effect of the presented number on the spatial attention of the subjects was found only in third-graders and in adults. This led to the conclusion that although 'children have an association between small numbers and "left" and between large numbers and "right" when they are as young as 7 years [...] [, they] have automatic access to magnitude information when perceiving Arabic numerals from 9 years of age' (van Galen & Reitsma, 2008, p. 109). An additional finding of the study was that with the increasing age, the SNARC effect is becoming increasingly categorical, indicating the

interaction between the SNARC and the numerical distance effect (Gevers, Verguts, Reynvoet, Caessens & Fias, 2006).

As already mentioned, the study of the spatial-numerical associations in children who have not yet received formal tutoring in elementary mathematics requires the adoption of special experimental paradigms, because such children do not yet possess the skills to perform the numerical tasks in which the SNARC effect manifests itself. Opfer and Thompson (2006) looked for the evidence of the existence of the directional biases in the preschool children using three tasks: serial counting (the HM task with explicit instructions regarding touching and vocalising), addition, and subtraction of an item from a set. They performed a cross-sectional analysis with the children grouped by age (3-, 4-, and 5-year-olds) and compared their performance with the adult participants. 98% of the children and all adults in the study counted a row of items from left to right. In addition, the proportion of the participants who both added items from left to right and subtracted from right to left increased with the age, reaching 50% for the 5-year-old group (around 70% of adults added and subtracted this way). The children with the strong left-to-right bias were older than those without it (4.64 versus 4.1 years). Further, the strength of the spatial bias was shown to correlate with the children's performance on the GN task — the children who showed the strongest bias performed well in the GN task across the whole tested number range (1–9), while the performance of the remaining children clearly suffered for numbers greater than 4 (cf. Le Corre & Carey, 2007). This provided the first instance of the evidence for spatial-numerical associations in children well before their exposure to reading.

In later work, Opfer et al. (2010) employed another experimental set-up in order to allow the investigation of the associations between numbers and space in young children. They used boxes with cards labelled both numerically and pictorially and the task of the children was to establish the correspondence between the cards using the numerical information only, ignoring the pictorial cues. The cards were numbered either in the left-to-right or right-to-left order. The children tested

in the experiment were pre-reading 4-year-olds. Opfer et al. found that in the spatial search task, the children were faster and more accurate when the cards were numbered from left to right. Also, while in the left-to-right condition the children tended to adopt the correct search strategy more and more often as the experiment progressed, this was not the case for the numbering in the opposite direction. Overall, based on the performance of the children on the task the authors concluded that ‘4-year-olds appeared to have robust expectations about verbal numbers increasing in a left-to-right order’ (Opfer et al., 2010, p. 765). In the same study Opfer et al. have also replicated their findings published earlier (Opfer & Thompson, 2006), this time obtaining the mean age of the children that show a strong left-to-right bias (in adding/subtracting items to/from a set) of 5.21 years (compared to the mean age 4.45 years of the children without such a bias). Later, Opfer and Furlong (2011) published the results of a further extended version of the study with more experimental conditions in the spatial search task. Taken all together, the data from the experiments of Opfer et al. suggest that the spatial-numerical associations may be established earlier than previously thought, that is when the children are acquiring the competence in the counting procedure, rather than when they start to be exposed to the biases present in the formal instruction in reading and mathematics.

In parallel, de Hevia and Spelke (2010) focused on the investigation if the predispositions to associate numerosity with space are present already in infancy. They conducted a series of four experiments with 8-month-old infants using the dishabituation paradigm and a combination of numerical and spatial visual stimuli. The numerical displays were ensembles of simple geometrical shapes varying in number, with the continuous parameters (such as the total surface area) controlled for. As spatial stimuli, centrally-positioned coloured rectangles with constant height and varying length were used. In the first experiment, the infants were familiarised with a sequence of numerical displays, either decreasing or increasing in number, and were tested on the sequences of spatial displays, both decreasing and increasing in length. It was found that the children looked longer at the test sequence with the

ordering that was opposite to the one they were habituated to, thus transferring the habituation from the numerical ordinal to the spatial ordinal domain. In the second experiment, an attempt was made to test if the children would be able to extract the positive association between the numerosity of a set and the length of a rectangle based on a small sample of examples. The infants were presented with combined stimuli (the set shown above the rectangle) in a randomised order, in which larger and larger sets were paired with longer and longer rectangles. The test displays contained novel numerosities and a corresponding rectangle, the relation between which either followed the positive number-length association or not. Here, the dishabituation did not occur, but the children looked statistically longer at the test stimuli congruent with the positive number-length association than on the incongruent ones. Although based on the logic of the habituation paradigm an opposite outcome would be expected, the result might be accepted as an evidence of the children's sensitivity to the positive association of number and length. The third experiment was analogous to the second one, the difference being that an inverse relationship between the numbers and length was established at the familiarisation phase (i.e. larger sets were paired with shorter rectangles). Neither habituation nor statistically significant preference for any of the test displays occurred. The final, fourth experiment, was aimed at testing whether the children simply had a baseline preference for the positive pairing of number and length. The stimuli from the experiments 2 and 3 were shown without the familiarisation phase, and the children's looking times were measured. The infants did indeed show a preference for the positive pairings over the inverse ones, but it was not large enough to account for the results obtained in the previous two experiments. Based on this, the authors concluded that in the experiments 2 and 3 the 'infants abstracted a specific positive relationship (but not an inverse relationship) between number and length from a small number of examples, and they generalised this relationship to new values' (de Hevia & Spelke, 2010, p. 658). Summarising, the results obtained by de Hevia and Spelke suggest that the tendency to associate numbers and space may be in-

nate, and that the association of larger numerosities with greater horizontal length is intrinsically preferred.

2.4 Numerical Knowledge in Humans

— Summary

The findings of the experimental studies on the development of human numerical skills, reviewed in sections 2.3.1 to 2.3.3, can be presented visually in form of a timeline. This is shown in figure 1.

Having presented the review of the essential findings related to the human numerical knowledge and its development, it is appropriate to indicate which of the mentioned experimental studies my embodied computational modelling experiments are going to focus on. These studies can be divided into two groups — the ones which results have driven some of the important design decisions made when formulating the models (and which therefore provide the basis on which the plausibility of the proposed models can be argued) and the ones which data were aimed to be reproduced (at least qualitatively) in the course of the simulation experiments. The findings belonging to the former group are:

- before children start to learn to count within a range of numbers, they are already able to recite a stable counting list within this range (Gelman, 1980; Wynn, 1992b; Le Corre et al., 2006);
- the spatial arrangement of the items being counted affects children’s counting accuracy (Beckwith & Restle, 1966; Potter & Levy, 1968), therefore to eliminate this variable, a fixed, simple arrangement is usually used, such as a row (Alibali & DiRusso, 1999; Graham, 1999; Le Corre et al., 2006);
- children’s spontaneous counting is characterised by a remarkably consistent, culturally-specific spatial bias (Tversky et al., 1991; Opfer et al., 2010);

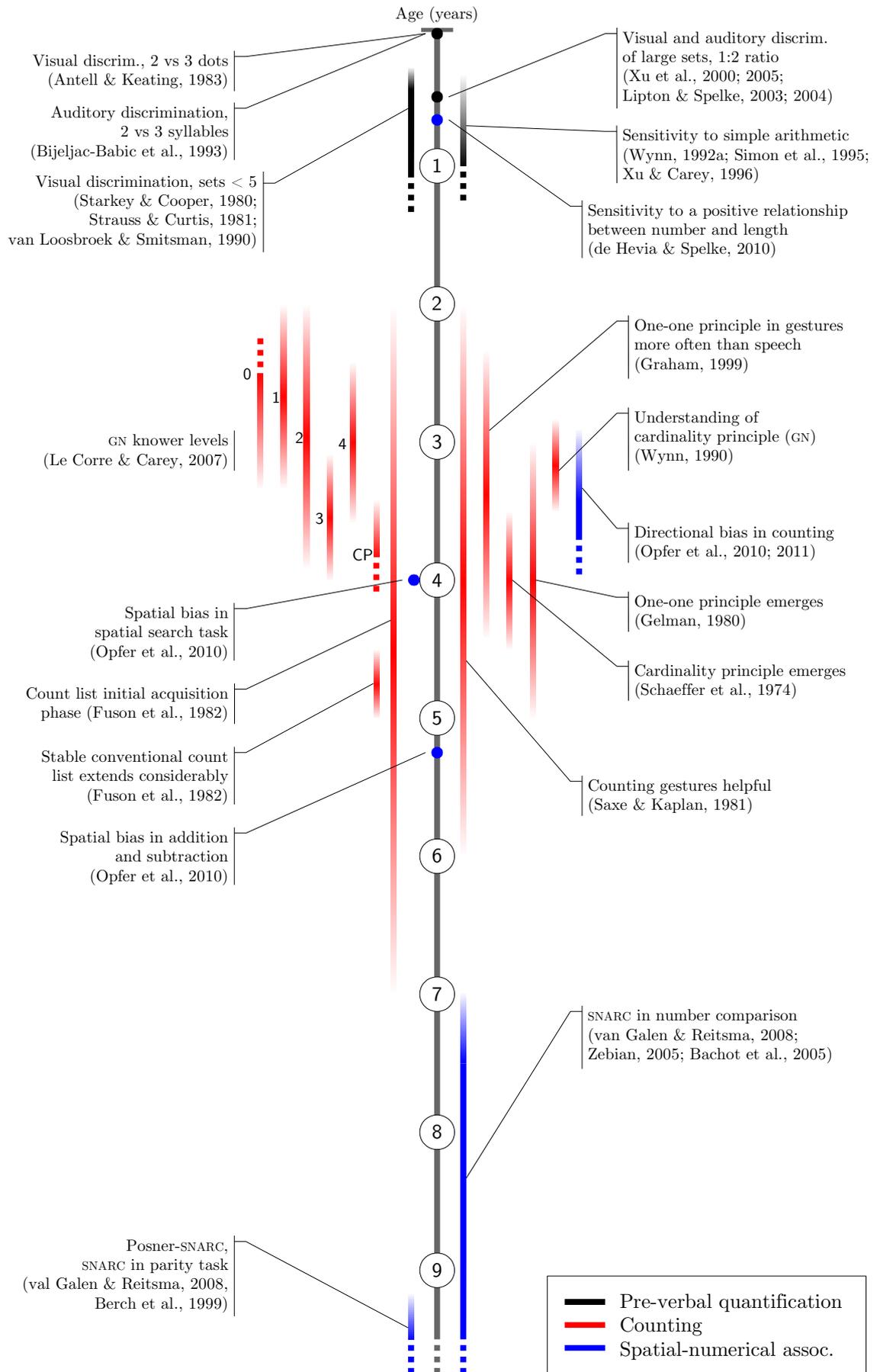


Figure 1: Timeline of human numerical knowledge development.

- the children’s adherence to the counting principles is evident in their gestures prior to speech (Graham, 1999);
- being able to execute the counting procedure correctly does not necessitate the ability to use it as a tool in task solving (in terms of success in the HM task, Schaeffer et al., 1974); furthermore, at least according to some researchers, even consistent success in the HM task does not imply possessing the genuine understanding of the counting principles (in terms of success in the GN task Wynn, 1990; Le Corre et al., 2006);
- spatial-numerical associations are most likely established before the onset of formal education in reading and mathematics (Opfer et al., 2010).

The experimental data against which the behaviour of the proposed models is going to be compared are:

- various characteristic features of the the children’s learning of the count list reported by Fuson et al. (1982);
- the effect of the presence and absence of the counting gestures on the children’s counting accuracy (Alibali & DiRusso, 1999);
- the effect of the size of the counted set on children’s counting accuracy (Alibali & DiRusso, 1999);
- the relative frequencies of the occurrence of specific counting errors in the children’s counting (Alibali & DiRusso, 1999);
- response time patterns characteristic for the number size and numerical distance effects in the number magnitude comparison task (Schwarz & Stein, 1998);
- response time patterns characteristic for the SNARC effect in the number parity judgement and number magnitude comparison task (Gevers, Verguts et al., 2006);

- response time patterns characteristic for the Posner-SNARC effect in the visual target detection task (Fischer et al., 2003).

The behavioural findings discussed in the present chapter provide the necessary background for the review of the attempts to understand more closely the mechanisms behind our numerical knowledge using the computational modelling methods, which is presented next.

Chapter 3

Computational Modelling in Mathematical Cognition

In this chapter I introduce the major computational modelling approaches for the human numerical skills introduced in chapter 2. Computational modelling is a useful tool in the study of cognition. As Mareschal (1998) puts it, it ‘provides a concrete vehicle with which to explore the interdependence of structural and processing constraints in a complex system and linking these to the behaviour that emerges from the system as a whole.’ (Mareschal, 1998, p. 148). In their own review of the computational modelling of numerical cognition, Zorzi et al. (2005) emphasise the way in which computer simulations supplement formulating theories verbally as follows: ‘In contrast to the loose formulation of traditional verbal theories, computational models need to be explicit in any implementational detail and can produce highly detailed simulations of human stimuli.’ (Zorzi et al., 2005, p. 67). In other words, the possibility to not only describe hypothetically but also to *simulate* a model of a cognitive capacity, allows it to be analysed with a much higher degree of scrutiny and to be assessed quantitatively. This, in turn, makes it possible to compare competing theories in a meaningful and fair way. Crucially, the capacity of the computational models is not limited to the reproduction of the behaviour. In addition, they can be used to generate fine-grained predictions about the novel situations, which can then be tested on humans.

As pointed out by Zorzi et al. (2005), several avenues in the area of modelling

of human mathematical abilities can be distinguished. In this chapter I focus solely on the models of the skills and phenomena that are the topic of my studies, that is non-verbal magnitude processing (section 3.2), counting (section 3.3), and the SNARC effect (section 3.4). This is preceded by a brief discussion of conceptual (i.e. non-computational) models of historical and theoretical importance (section 3.1). To the best of my knowledge, up to the time of writing there have been no studies that employed the embodied robotics methodology to model numerical skills, therefore my review mentions only ‘disembodied’ computational models. However, because — as discussed in chapter 2 — gestures are an important aspect of learning to count, and because there is a considerable amount of research on gestures conducted in the robotics field, at the end of this chapter (section 3.5) I provide an overview of a few selected robotic experiments that focused on gestures, even though they were not concerned with the modelling of mathematical cognition.

3.1 Qualitative Models of Human Mathematical Skills

I start the overview of the past modelling studies in mathematical cognition with a brief discussion of selected qualitative models. In most cases these models have never been used for simulations, what would allow a systematic exploration of their actual agreement with experimental data or of the predictions they produce. These models are however widely cited and they influenced later computational modelling attempts — and for that reason deserve mentioning.

The *accumulator model* was originally proposed in the context of animal skills in timing tasks (Gibbon, 1981) and, by extension, it has been applied to explain animal sensitivity to numbers (Meck & Church, 1983; Church & Meck, 1984; Church & Broadbent, 1990). The main components of the model are the *pacemaker*, the *mode switch* and the *accumulator*. The pacemaker is assumed to generate pulses at a constant rate. The mode switch controls the passing of the pulses to the accumulator

which aggregates all pulses received. The value of the accumulator can be stored and retrieved from the memory allowing further processing, for example comparison. Finally, the mode switch may operate in various modes (run/stop or event mode) which allow adopting the model to different tasks, like the perception of duration, rate or numerosity. Two important properties of the accumulator are the continuity of the representation and its stochasticity according to the Weber's law (compare section 2.1.1).

An inexact nonverbal quantification system based on the accumulator model has been suggested to underlie more precise numerical abilities that utilise symbolic representations of numbers in humans (Gallistel & Gelman, 1992, 2000). More specifically, Gallistel and Gelman proposed that the properties of the accumulator system provide the basis on which verbal counting and arithmetic reasoning are founded. The acquisition of the meaning of number symbols was hypothesised to correspond to building a bi-directional mapping between the discrete symbols and the corresponding continuous and imprecise values of the accumulator. Appropriate conversion between these formats was the preliminary step of more complex mathematical tasks like number fact retrieval. Over the years various extensions to the model have been proposed, including the relaxation of the assumption of the inherently sequential nature of the accumulation process (Cordes & Gelman, 2005), as well as the additional innate representations consisting of the integer concept of unity and a successor function, which were proposed to form the basis of the natural number concept (Leslie, Gelman & Gallistel, 2008). The accumulator model is one of the longest existing qualitative models of nonverbal magnitude representation and different computational models can be quoted which share at least some of its properties (see for example Dehaene, 2001; Grossberg & Repin, 2003).

In the last two decades of the 20th century various high-level models of the overall architecture of the cognitive arithmetic system have been put forward. All these models attempt to encompass the multimodal and/or multi-format aspect of numbers. They refer to relatively complex cognitive tasks and explain, albeit

only qualitatively, human behavioural data. McCloskey and colleagues proposed a modular model of numerical processing and calculation (McCloskey, Caramazza & Basili, 1985; McCloskey, 1992; McCloskey & Macaruso, 1995), which separated the processes of number comprehension and production in various formats (e.g. Arabic digits, written number words) from each other and from the calculation mechanisms. At the core of the model was an abstract and amodal internal representation of number. In contrast, the *encoding complex hypothesis* (J. M. Clark & Campbell, 1991; Campbell & Clark, 1992; Campbell, 1994) advocated specific number codes for various formats of number as well as corresponding different calculation processes. Multiple number codes were assumed to be interconnected in one associative network, constituting the multi-component (but not modular) ‘encoding complex’. The *network retrieval model* (M. Ashcraft, 1987; M. H. Ashcraft, 1992) focused on the retrieval of mathematical facts, such as results of addition or multiplication operations, from long-term memory. The crucial structural aspect of the model was reliance on the strengths of connections between nodes representing the number facts. These strengths reflected varying degrees of relatedness among the nodes and were intended to account for variations in retrieval times. Based on the observation that in humans the time needed to recall information depends on experience, the strength of the connections was assumed to be a function of the frequency of occurrence of particular mathematical facts in the early education. Finally, the *triple-code model* (Dehaene, 1992; Dehaene & Cohen, 1995) was based on two assumptions. First, and similarly to the model of Campbell and Clark, numbers in the triple-code model were represented mentally in three different codes (namely auditory verbal, visual Arabic and analogue magnitude) and each of these codes was connected with the dedicated mechanisms of input and output, as in the model of McCloskey. Second, and in contrast with both older models, in the triple-code model specific numerical tasks were tied to a specific number code, rather than to all codes or to a central amodal representation. Notably, one of the codes of the triple-code model was assumed to correspond to the hypothetical nonverbal approx-

imate quantification system shared by animals and humans (Gallistel & Gelman, 1992; Feigenson et al., 2004).

3.2 Models of Magnitude Representation and Elementary Numerical Abilities

Increasing availability and popularity of computers as well as the progress in areas related to cognitive modelling (such as artificial neural network frameworks), allowed these tools to be exploited for the benefit of the study of cognition. This resulted in the appearance of the first quantitative models of numerical cognition in the early 1990s.

3.2.1 The First Quantitative Model (Dehaene & Changeux, 1993)

The study published by Dehaene and Changeux (1993) is among the first efforts to quantitatively model the development of elementary numerical abilities possessed by animals and humans using a formal connectionist architecture. It focused on the perception and simple processing of non-verbal visual and auditory stimuli. The architecture of the model was modular and to a large extent hand-wired. At the heart of the model was a cluster of units called *numerosity detectors* which were designed to respond selectively to the numerosity of the stimuli presented at the input to the system. The visual input was modelled as a 1-dimensional *retina* at which objects of various sizes could be placed, represented as Gaussian distributions. The retina projected into a further cluster of units which task was to dissociate the locations of objects from their size. This was achieved through an array of difference-of-Gaussian filters. Since in such an array each object is represented by a fixed pool of units regardless of its size, the numerosity of the visual stimuli was extracted simply by summing up the total activation of units over this array. This was done by another group of units, the *summation cluster*, realised with the use

of increasing thresholds. These units projected onto the final layer of numerosity detectors, which responded selectively to specific ranges of values of the total activity in the summation cluster. The auditory input was simulated with the use of short-term echoic memory also connected with the summation cluster. The activations in the numerosity detectors layer were used to simulate simple numerical tasks, like number discrimination or same-different and larger-smaller comparison tasks. The network was taught to perform these tasks using reinforcement learning. After the training the model exhibited the size and distance effects characteristic of animal and human performance. In the most elaborate instantiation of the model found in the paper it was shown how the system can autonomously ‘discover’ the larger-smaller relation between numerosities through unsupervised experimentation with addition and subtraction.

It is important to point out that the model of Dehaene and Changeux (1993) used a parallel process to extract the approximate numerosity of the stimuli. Therefore it demonstrates that accounting for the quantification skills of animals and human infants does not necessarily require assuming that they can count, that is serially enumerate items using a sequential process (see Gallistel & Gelman, 1992). As a pioneering effort in connectionist modelling of mathematical cognition it is deservedly one of the most widely cited works in this context. The fact that the connection weights in the central modules of the model of Dehaene and Changeux were hand-wired makes the model unsuitable for studying the development of the corresponding mechanisms and some interpreted this as an implicit assumption of innateness (Ahmad, Casey & Bale, 2002; Zorzi et al., 2005).

3.2.2 The Mental Representation of Magnitude

One of the longest standing questions in mathematical cognition is that of the nature of the mental code for numbers or, more generally, magnitude (Dehaene, 2003). In order to account for the various experimental phenomena connected with magnitude representation (see section 2.1.1) different computational approaches have been pro-

posed. Most of them are variations of the *number line* concept, in which magnitude is represented by a cluster of neurons dedicated for that purpose (for review see Zorzi et al., 2005, pp. 72–80). Two important parameters of such coding are the method of scaling and the character of variability. The scaling is most often assumed to be either linear (where the neuron index corresponds linearly to the represented magnitude) or compressed, usually in a logarithmic fashion. In turn, the variability may be constant regardless of the magnitude being represented or scalar, that is increasing along with it. These two properties of the magnitude representation system have been often assumed to be the source of the number size and numerical distance effects. The accumulator model (Gallistel & Gelman, 1992) is an example of magnitude coding with linear scaling and scalar variability while the numerosity detectors in the model of Dehaene and Changeux (1993) exhibit logarithmic scaling.

Despite the ample experimental and modelling data the ‘linear versus logarithmic’ controversy is still not resolved entirely. One of the major problems is the non-obvious ambiguity of the experimental results. For instance Brannon, Wusthoff, Gallistel and Gibbon (2001) have found evidence for the linear encoding of numbers in a bird brain, in a study constructed using the *number-left* paradigm where pigeons were required to compare a constant reference number with the number remaining after subtraction. Their results falsified the predictions of the logarithmic magnitude encoding and suggested linear scaling with scalar variability. This was true however only under the assumption that the birds actually solved the task with the use of arithmetic. Dehaene (2001) simulated the experiments of Brannon et al. using a simple neural network model. He showed that the task did not require the pigeons to subtract at all, and could be solved simply by associating the numbers with the appropriate answers. When the task was solved this way, both schemes of representation tested by Dehaene — linear and logarithmic — lead to the identical behaviour, which was at the same time consistent with what Brannon et al. observed.

Another approach to magnitude representation found in the literature is the *numerosity code* (Zorzi & Butterworth, 1997, 1999; Zorzi et al., 2005), which utilises

linear scaling without variability. The magnitude is encoded by the number of activated units in the cluster, in a thermometer-like fashion. Zorzi and Butterworth (1999) demonstrated that, in contrast to what has been often assumed, neither analogue code nor variability in the representation are necessary to reproduce the behavioural effects found in the number comparison task and that these effects may as well appear at the decision stage. In later work, Stoianov, Zorzi, Becker and Umiltà (2002) compared the numerosity code with other commonly used magnitude codes and found that the learning of arithmetic facts using the former required less epochs of training than any of the latter. Furthermore, out of all tested codes only the numerosity code allowed for reproducing the effects related with response times which are found in humans, namely the problem-size effect (Zbrodoff & Logan, 2005).

Summarising, despite a lot has been learnt about the matter, the question of the mental representation of magnitude remains to a large extent unanswered. All proposed models account for some behavioural data, often with significant overlap with each other. More insight on the issue is likely to be shed based on the data from neuroscience. For example, direct single-cell recordings in the prefrontal cortex of the monkey brain confirmed the existence of neurons which respond selectively to specific numerosities (Nieder, Freedman & Miller, 2002). The properties of the neurons discovered by Nieder et al. were in favour of the compressed coding rather than the one with scalar variability (Nieder & Miller, 2003). Strictly speaking however, multiple types of coding could well co-exist within the brain (Brannon et al., 2001), and some codes may be prerequisite for developing other ones (Zorzi et al., 2005). Note for example that in the model of Dehaene and Changeux (1993) one can identify both the thermometer-like numerosity code (in the summation cluster) and the classical number line code (in the numerosity detectors).

3.2.3 Between Subitising and Counting

The computational model of Peterson and Simon (2000) focused on the human ability to subitise, that is to immediately apprehend the numerosity of a small visually presented set without reverting to serial enumeration (Kaufman et al., 1949). In particular they proposed two models with the aim of investigating the possible reasons behind the existence of a specific upper limit for the number of items that humans are able to subitise (believed to be around 4). The first model was formulated within the ACT-R framework (Anderson, 1993). It implemented two strategies to report the number of items in a set: one based on pattern recognition and one based on sequential enumeration. The first strategy was realised by storing all encountered configurations of objects in the memory and recalling the result of the enumeration upon encountering the same arrangement of objects again. The standard mechanisms present in ACT-R were used to model such aspects of the memory trace as its decay with time, its strengthening upon multiple exposure, as well as the impact of its strength on the retrieval time. The second strategy, the counting procedure¹, was ‘an implementation of the counting principles described by Gelman and Gallistel (1978)’ (Peterson & Simon, 2000, p. 120). Presented with a set to be enumerated, the model used the recognition strategy whenever possible (i.e. if the configuration of objects has been seen by the model before and its memory trace was strong enough) and reverted to the serial enumeration otherwise. The simulations with the ACT-R model indicated its strong reliance on recognition for numerosities lower than four, its primary reliance on enumeration for numbers greater than four, and an intermediate situation for the number four itself. In addition, the chronometric data obtained using the latency measurement mechanisms available in ACT-R reflected to some degree the pattern of the response times found in humans. This provided evidence that the upper limit for the ability to subitise may emerge as

¹Since Peterson and Simon (2000) admit themselves that in their model ‘the counting procedure is not intended as a detailed model of the information processing that takes place during such enumeration’ (see Peterson & Simon, 2000, p. 103 and note 2 on p. 120) and indeed they do not provide sufficient details about how the enumeration procedure has been implemented, their model is not considered together with the other models of counting in section 3.3.

a result of the combinatorial characteristics of visually presented sets of items — since the number of different spatial arrangements of a set grows very fast with its cardinality, for larger numbers it is increasingly difficult to rely on the pattern recognition strategy. The second model proposed by Peterson and Simon was a 3-layer feed-forward neural network trained by backpropagation. The network accepted a binary vector representing the arrangement of the objects on a 2-dimensional retina as input and produced at the output a symbolic representation of numerosity based on a one-hot code. In this model, only subitising was simulated. The results of the simulations of the second model were somewhat less clear-cut than those of the ACT-R model. While the numerosities below the hypothesised subitising limit (i.e. lower than 4) were always learnt easily, for the larger numerosities (5–8) the findings were not consistent. Depending on the choice of the parameters of the model (the number of hidden units or the problem size), different results were obtained. Crucially, the model often exhibited the same behaviour for some of the larger numerosities as for the smallest ones, which is in conflict with the subitising hypothesis.

A modular connectionist model which, similarly to the ACT-R model of Peterson and Simon (2000), incorporated separate routes for subitising and counting has been proposed by Ahmad and Bale (2001). The modelled task was to report the number of objects in a visual scene in the absence or in the presence of time constraints. The model was realised in the mixture-of-experts architecture with two sub-modules implementing the two quantification strategies. At the highest level, a gating network decided which of the two strategies was used in the particular instance, depending on whether the time constraint was imposed or not.

Although the original paper reports some preliminary results of the simulation of the model as a whole (see Ahmad & Bale, 2001, p. 87), in their further work the same authors present the two pathways of the model in more detail and as separate models (Ahmad et al., 2002). Unfortunately, the study of how the allocation of the two strategies to the specific sub-domains of the task (subitising for small numbers and counting for larger ones) could self-organise in the mixture-of-experts architec-

ture, which could provide additional insight into the sources of the subitising limit as investigated by Peterson and Simon (2000), was not followed up. Accordingly, the two pathways of the model of Ahmad and Bale (2001) will be considered as separate models, as described by Ahmad et al. (2002). The subitising model is briefly reviewed below. The model of counting in turn is discussed in detail in section 3.3.5 along with the other models of this skill.

The model of subitising proposed by Ahmad et al. (2002) consisted of a number of modules connected in a sequence. The input to the model was a two-dimensional retina with 36×18 units. The retina was divided into non-overlapping receptive fields of 3×3 units on which objects could be located. The first layer of the model was a *second-order network* which produced a size-invariant representation of the visual scene. It consisted of 72 units (one for every receptive field) which indicated the presence of an object in the corresponding receptive field. The second module, the *weight sharing network*, provided a translation-invariant representation of the visual scene. Although specific details are not provided, it can be deduced that this 15-unit layer implemented a summation coding scheme with the number of units activated in the layer proportional to the number of objects in the visual scene, a solution similar to those employed in the previous studies (Dehaene & Changeux, 1993; Zorzi & Butterworth, 1997). Activations in this layer were ‘padded to 36 dimensions’ (Ahmad et al., 2002, Table 4) and used as the input to the *magnitude representation* module, realised as a 1-dimensional 36-unit Self-Organising Map (SOM, Kohonen, 1982). The explanation how ‘padding’ was done and the motivations for such processing were however not provided. Finally, the activations of the magnitude representation module projected onto the output module, which was a feature map-based *verbal representation* of the number words. It was realised as a SOM as well, with rectangular 8×8 topology.

The aforementioned components of the model were all trained separately using a variety of both supervised and unsupervised training regimes. Importantly, the magnitude representation module was trained using the unsupervised SOM train-

ing algorithm, in contrast to most of the previous works in which the magnitude representation was wired by hand. As a result, the topological population code representing the magnitude, rather than being imposed, emerged from the properties of the representation available in the previous layer (the summation coding network). The same approach has been used to form the verbal module. Here the input to the map was constructed as a 16-element ‘phonetic’ representation of number words. Ahmad et al. (2002) do not provide specific details about how the number words were actually decomposed, apart from stating that ‘each element within the input vector represented a phoneme needed for all 22 numbers’ (Ahmad et al., 2002, p. 183). In any case, one can assume that after the training has been completed, the 64-unit map represented the considered number words topologically with areas corresponding to similarly-sounding words located near each other. After the magnitude and verbal maps were fully developed, a mapping between them was constructed, allowing for (in principle) bi-directional conversion between the representations. This mapping was trained in a supervised fashion, by strengthening the links between the areas in the two maps which co-activated when a number word together with a corresponding visual scene were presented to the model, according to the Hebbian learning rule.

The performance of the fully developed model was compared with the selected previous modelling approaches (namely Dehaene & Changeux, 1993; Peterson & Simon, 2000), pointing out the similarities and differences in obtained results. The representation of the magnitude which formed in the model during the training was assessed qualitatively for having a potential to exhibit the size and distance effects (Ahmad et al., 2002, p. 182). However, the actual simulations which would demonstrate the presence of these effects in the behaviour of the model were not conducted.

3.2.4 The Temporal Structure of Numerical Processing (Grossberg & Repin, 2003)

Grossberg and Repin (2003) were among the first to formulate an artificial neural network model of the dynamical structure of neural number processing in simple numerical tasks. Their paper actually describes two sets of simulation experiments, the second one based on an extended version of the model formulated in the first one. The basic model, dubbed *Spatial Number Network*, was aimed at explaining a subset of behavioural data on non-verbal magnitude recognition and comparison, and consisted of three components. The *preprocessor* handled the conversion of sensory inputs coming from different modalities (visual or auditory) into a neural activation roughly proportional to the numerosity of the stimuli, in a fashion similar to the accumulator model operating in the event mode (Meck & Church, 1983). The signal from the preprocessor activated the *spatial number map* which implemented a linearly-oriented number representation scheme with the variability increasing along with the magnitude. The final layer of the model, the *comparison wave*, consisted of direction-sensitive units which detected the direction of reorganisation of the activation of the spatial number map upon the input stimuli change and thus enabled the model to perform magnitude comparison. These parts of the model were hand-crafted in order to qualitatively account for the target behavioural data. Grossberg and Repin argued that these were ‘specialised combinations of mechanisms that have previously been used to model processes of motion perception, spatial attention and target tracking’ in their previous works (Grossberg & Repin, 2003, p. 1112). The free parameters of the model were adjusted (partially by optimisation and partially by trial and error) in order to obtain a quantitative fit to the experimental data. The model reproduced, although not with perfect accuracy, a wide range of behavioural data connected with number recognition — response distributions found in rats (Mechner, 1958) and human number reading times measured using gaze duration (Gielen, Brysbaert & Dhondt, 1991) — and comparison — animal and human error rates (Washburn & Rumbaugh, 1991; Dehaene et al., 1990), semantic priming effects

(Brybaert, 1995) and number size and numerical distance effects (Parkman, 1971; Link, 1990).

In the second part of their paper, Grossberg and Repin (2003) describe the *Extended Spatial Number Network* model, aimed at demonstrating that the basis for a place-value number system, typical for symbolic numerical notations used by humans, may emerge from the associative processing between the *where* and *what* streams in the brain. This extended model employed a two-dimensional spatial number map, which, together with a training regime in which single-digit numbers were presented prior to two-digit ones, allowed for the formation of parallel associative links from multiple linguistic categories (e.g. tens and units) to the corresponding areas of the ‘primary’ spatial number map. These links were responsible for the induction of multiple, parallel comparison waves along one of the dimensions of the two-dimensional spatial number map, which were then aggregated in order to implement multi-digit number comparison. Both error rates and chronometric data for this task were simulated. Most prominently, the interactions between the multiple comparison waves enabled the model to account for a paradoxical reverse numerical distance effect found in human response times across the boundaries of decades (Brybaert, 1995). However, as the authors themselves admitted (Grossberg & Repin, 2003, p. 1138), finding the values of the many model parameters which would give a reasonably good fit to the data was not an easy task.

In the context of the work presented in this thesis it is important to highlight that although the models of Grossberg and Repin (2003) contain elements which suggest counting (more specifically, the preprocessor), this process was not modelled in their study. The correct sequential enumeration that would lead to counting (in the visual modality) was assumed to take place externally to the model (Grossberg & Repin, 2003, p. 1112). Furthermore, depending on the simulated phenomenon, either actual sequential input to the preprocessor was used or this part has been replaced by a one-off stimulus fed directly to the spatial number map (Grossberg & Repin, 2003, p. 1116). Finally, the response readout occurred through different

pathways depending on the task (Grossberg & Repin, 2003, p. 1122) and the error rates of the model were not measured literally, but were assumed to be a function of the difference between the activation values of the model outputs (Grossberg & Repin, 2003, p. 1118).

3.2.5 Identifying the Sources of the Behavioural Effects

A consistent path of progressive modelling of the elementary numerical skills can be found in a series of papers by Verguts and collaborators (see Verguts & Fias, 2008, for review). The first model (Verguts & Fias, 2004) focused on the numerosity encoding schemes believed to be employed in the brain, especially in the light of the recent findings in neuroscience which confirmed the existence of neurons able to act like numerosity detectors (Nieder et al., 2002; Nieder & Miller, 2003). Verguts and Fias addressed the questions about the nature of the input required for the numerosity detectors, whether this type of coding has to be hard-coded and also explored the possible relations of this representation system with symbol manipulation capabilities. In the first set of simulations they showed that when a 3-layer feed-forward artificial neural network is trained to implement numerosity detection in its output layer, the units in the hidden layer robustly self-organise into a summation-coding layer. This provided evidence that summation coding may be an efficient intermediate step to obtain number-selective coding (see also Zorzi et al., 2005, pp. 80–81). Then, using unsupervised learning techniques, they demonstrated that number-sensitive neurons exhibiting properties consistent with those found by Nieder and Miller in vivo, may spontaneously self-organise based on summation coding input. In contrast to what has been suggested previously (Dehaene, 2002) this result implies that, since they are easily learnt, neural number detection systems do not have to be necessarily innate. In the final simulation Verguts and Fias found that numerosity detectors which self-organised in response to the non-symbolic numerical stimuli can be later associated with symbolic stimuli and that the latter input format leads to obtaining more precise representations than the former. The

fact that the symbolic representations may thus be grounded in non-symbolic ones provided a possible account for the reason why the processing of symbolic stimuli (such as Arabic numbers) in humans is subject to the effects suggesting the fuzziness of the representation, like the numerical distance effect, despite the orthogonal nature of the symbols.

In further work Verguts, Fias and Stevens (2005) employed a number representation system, with properties suggested by the previous study (Verguts & Fias, 2004), to address the issue of the source of the number size and numerical distance effects. The main problem addressed by Verguts et al. was that the predictions that may be formulated based on the previously proposed accounts for the size and distance effects — compressed scaling (Dehaene, 2003), scalar variability (Gallistel & Gelman, 1992), and magnitude coding (Zorzi & Butterworth, 1999) — were inconsistent with the experimental findings reporting symmetric priming patterns in number naming and parity judgement as well as with the lack of the number size effect in these tasks (Dehaene et al., 1993; Fias et al., 1996; Reynvoet & Brysbaert, 1999; Butterworth, Zorzi, Girelli & Jonckheere, 2001; Reynvoet, Brysbaert & Fias, 2002; Reynvoet, Caessens & Brysbaert, 2002; Reynvoet & Brysbaert, 2004). The semantic number representation used by the model proposed by Verguts et al. utilised a place-coding system with linear scaling and constant variability. The input to the model was assumed to correspond to the symbolic Arabic notation and was implemented using orthogonal, one-hot coding. Weights between the input and the semantic representation layers were hand-wired to yield coding with the desired properties. For each of the simulated tasks (naming, parity judgement and comparison) the model was extended with a dedicated output layer consisting of an appropriate number of units. Also, in order to implement the number comparison task, two parallel input and semantic layers were used. The model was then trained to perform the tasks in a supervised fashion using the Widrow-Hoff delta rule (Widrow & Lehr, 1990). The frequencies of the presentation of numbers during training were biased according to the frequencies of their occurrence in language in daily use as reported by Dehaene

and Mehler (1992). Similarly to Grossberg and Repin (2003), Verguts et al. used a continuous-time neural network implementation, which permits the direct modelling of the response times (see chapter 4). In simulations, the model reproduced the human pattern of behaviour, i.e. it exhibited the number size and numerical distance effects in the number comparison task. It did not show the size effect in the parity judgement and naming tasks, while exhibiting in these tasks a symmetric priming pattern with respect to the absolute numerical distance between the prime and the target. This suggested that the size and distance effects in the comparison task should be primarily attributed to the decision stage rather than to the representation stage (see also Zorzi & Butterworth, 1999). More specifically, Verguts et al. identified the compressive pattern of the connection weights between the semantic and output layers, obtained as a result of the number frequency manipulation during training, to be the source of the aforementioned effects in the comparison task. In contrast, in the other two tasks the priming distance effect was caused by the fuzziness of the semantic representation. Since this fuzziness in the proposed model was symmetric, the obtained response time patterns were symmetric as well. The model of Verguts et al. served as a basis for the later models of the SNARC effect published by the same group, which are discussed in section 3.4.

3.2.6 Grounding Quantification in Perception

Rajapakse, Cangelosi, Coventry, Newstead and Bacon (2005b, 2005a) studied the relations between the symbolic and non-symbolic aspects of the human numerical abilities from a slightly different perspective. They considered how human numerosity judgements and the use of the linguistic quantifiers can be grounded in perception. Reproducing an experimental set-up used in a behavioural study conducted in parallel (Coventry, Cangelosi, Newstead, Bacon & Rajapakse, 2005), their model was presented with visual stimuli containing striped and white fish. The task of the model was to produce the acceptability ratings for five linguistic quantifiers (‘a few’, ‘few’, ‘several’, ‘many’, and ‘lots’) based on the number of striped fish in the

scene, with the white fish acting as distractors. The model of Rajapakse et al. consisted of 4 modules. The first two, the *vision module* and the *compression networks* performed dimensionality reduction and produced a compact representation of the input visual scene. The *quantification network* was trained to produce the numerosity judgements for both types of fish in the scene. Finally, based on the input from the compression networks and the quantification network, the *dual-route* network performed language production (that is it produced the acceptability ratings for the five considered quantifiers) and constructed a ‘mental image’ of the original visual stimuli. The distinct features of the model of Rajapakse et al. were that it was fed with real images as the visual input instead of using a simplified abstract representation of the visual scene, it focused on the use of vague linguistic quantifiers rather than number words which have a precise semantic meaning, and that the effects of the context on the use of these quantifiers were considered.

3.3 Models of Counting

In the present section I review the existing models of counting. Not many models have been published to date, and most of them focus on a very narrow aspect of this complex skill. The most thorough attempt to model counting has been published by Ahmad et al. (2002), and therefore it is given the most prominent attention.

3.3.1 Programming a Neural Network to Count Identical Sequential Stimuli (Amit, 1988)

Amit (1988) proposed a neural network that counts identical stimuli occurring in a temporal sequence (e.g. clock chimes). The model was based on an extended version of the Hopfield network (Hopfield, 1982) able to store and recall temporal sequences by employing a set of additional time-delayed synapses. Upon receiving a stimulus, such a network, rather than immediately converging to a single stable state as a typical Hopfield network, follows a sequence of transient but well defined

states (quasi-attractors), spontaneously transitioning from one state to the next. The idea was to adjust the parameters of the network in such a way that the transition between the quasi-attractive states would no longer be spontaneous, but would require a presence of an external stimulus in order to occur. In effect, the dynamics of the obtained process would implement counting of the external stimuli. The transient states of the network would represent numbers. The output of the network — the result of counting — would be read out as the network state after all stimuli that were to be counted have arrived. Amit provided the equations using which the synaptic weights of connections in the network that would lead to the desired dynamics could be computed given the previously determined sequence of the ‘number states’. He also described the results of two simulations which demonstrated the functioning of the system under the low and high level of the memory load.

It is important to point out that one of the consequences of the Amit’s approach is the assumption that the ability to recite the correct sequence of number words is prerequisite for counting (Amit, 1988, p. 2143). The paper is not concerned however with the modelling of the process of learning to count. Using the author’s own words, it is ‘basically an exercise in connectionist programming’ (Amit, 1988, p. 2143), since all parameters of the model were hand-wired in order to achieve the desired behaviour. The goal of the study was to demonstrate that artificial neural networks possess the capability to enumerate sequential stimuli despite them being identical, what apparently has been questioned at the time (see Amit, 1988, p. 2142).

3.3.2 Learning the Sequence of Number Words (Ma & Hirai, 1989)

Some of the aspects of the development of counting have been addressed by Ma and Hirai (1989), whose study focused on the children’s learning of the sequence of the number words. Ma and Hirai considered three specific phenomena which characterise the use of the counting list as its learning progresses (Fuson et al., 1982, see

section 2.3.2): the division of the learnt list into three distinct portions (stable conventional, stable non-conventional, and unstable), special difficulties with learning the irregular number words (e.g. ‘fifteen’), and the difficulties with recalling the next number word from the middle of the counting list when only a single preceding word is given. The model of Ma and Hirai was based on a combination of a heteroassociative network with a recurrent inhibitory network, and was trained in a supervised fashion. The central assumption was that the counting list is learnt by associating the number words with each other. At the beginning, multiple associations between the numbers are assumed to exist (e.g. 1 can be associated with both 2 and 3), and the observed errors result from the competition between the associations. Along with the progress of the learning, the ‘correct’ associations become stronger with respect to the ‘incorrect’ ones, and as a result the errors occur more and more rarely. Three sets of simulations showed that the model is able to exhibit the behaviours qualitatively similar to the considered effects found in children. Each of the three effects was however demonstrated in a slightly different experimental set-up, and in some cases the modifications to the training regime, like explicitly dividing the counting list into subsequences learnt separately, were necessary.

3.3.3 Artificial Neural Network Architectures for Counting (Hoekstra, 1992)

Hoekstra (1992) compared the performance of three artificial neural networks architectures in the counting task. The task was defined in a similar way as considered by Amit (1988), and was to report the number of sequential pulses incoming at the network input until a time step with no pulse has occurred, on which the output of the network should reset to zero. The three architectures considered by Hoekstra were:

- the *time-delay network*, in which temporal processing is achieved by assuming that the consecutive input units represent input occurring at the consecutive time steps;

- the *Elman network* (Elman, 1990) which is capable of literally temporal processing achieved using the recurrent connections in the hidden layer with the delay of 1 time step;
- a *mixed network*, which was a simple extension to the Elman network proposed by Hoekstra, with additional links with the delay of 1 time step from the input node to the hidden layer.

In the comparison of the training efficiency of the three networks, the Elman network proved to be the slowest to converge, and the time-delay network the fastest. The main finding of the paper was that the proposed mixed network achieved the performance very close to that of the time-delay network, while accepting the input in a serial form rather than parallel, thus offering both the literal temporal processing and the high training performance. Finally, Hoekstra demonstrated that the mixed network can be trained to count up to 16 items.

3.3.4 Spontaneous Counting in an Elman Network (Rodriguez, Wiles & Elman, 1999)

Rodriguez et al. (1999) conducted a study on training a recurrent neural network to count using a formal language framework. They trained a small Elman network (Elman, 1990) with 2 inputs, 2 outputs and 2 hidden units to recognise the structure of a deterministic context-free language of the form $a^n b^n$. The task of the network was to predict the next input character in the string and in order to do this accurately, the network had to count the number of a 's and b 's presented at the input. Despite the task being tough to learn (the success was achieved in 16% of the trials), Rodriguez et al. obtained the networks capable of performing the task and also, to a certain extent, capable of generalising. The mechanism used by the successful models to solve the task was particularly interesting, especially when compared to the traditionally employed computational approaches to represent the numerosity in the connectionist architectures which were discussed in section 3.2. In short, the

two hidden units of the successful Elman networks formed a discrete-time dynamical system with properties dependent on the current input to the network. During each of the two parts of the input sequence (a 's or b 's) the activity of the hidden units followed a trajectory determined by the current properties of the system. The transition from a to b at the input caused a change in the dynamics of the system and, consequently, made the hidden units follow a second trajectory, complementary to the first one. Provided that the coordination between the two trajectories was accurate enough, the hidden units state passed the decision boundary after making the same number of steps along both trajectories, what yielded correct predictions. Having analysed the nature of the mechanism obtained through training, Rodriguez et al. have shown analytically that, in theory, it is possible to hand-wire the system to be able to process strings of arbitrary length. In a follow-up work, Boden, Wiles, Tonkes and Blair (1999) have analysed why learning the task in the chosen framework was so hard. The complex nature of the error surface, in which the solution which provides the required network dynamics lies near a bifurcation point (where small changes in the network weights induce drastic changes in the network behaviour), was identified to be the reason.

The study of Rodriguez et al. (1999) showed that the Elman network can develop the ability to count the symbols in the input sequence using a non-trivial mechanism as a result of supervised learning with backpropagation through time. However, theirs was not a cognitive model of the human ability to count, and no comparisons with psychological data were made. The model can be seen as an example of preverbal counting, since no explicit external number word sequence has been employed in the experimental set-up (Ahmad et al., 2002). It can be argued however that the internal states and the interactions between them which emerge in the course of training constitute an internal 'vocabulary' which the model develops and uses to solve the task, leaving this question open to interpretation.

3.3.5 Multi-Net Simulation of Counting (Ahmad et al., 2002)

A sophisticated connectionist architecture for counting has been proposed by Ahmad et al. (2002). On the top level of the architectural design the model consists of three modules connected in a sequence: the *mapping module* which parses the visual scene, the actual *counting module*, and the *output module*, which maps the output of the counting module onto the final model response.

The initial processing stages in the model are identical to those employed in the subitising model described in the same paper and discussed earlier (Ahmad et al., 2002, pp. 177–187, see section 3.2.3), up to, and including, the scale invariant network. Since in the context of counting the spatial arrangement of the objects is important, the mapping module does not contain the translation-invariant layer. Instead, another kind of processing takes place which unfortunately is described in a confounded way and in insufficient detail. Quoting the authors verbatim:

Since the scale invariant visual scene outputs objects within a receptive field, with no gaps between objects, this was modified before presentation to the counting module to include gaps at appropriate points. The output of the mapping module was further reduced in dimension to match the total number of objects that the counting route was capable of detecting. (Ahmad et al., 2002, pp. 188–189)

Visual scene is converted from a 72-dimensional vector to a 44-dimensional vector with spaces between objects. (Ahmad et al., 2002, Table 7)

Later in the paper, it is added:

On separate runs of SCOUSYST the sizes of the layers of the network were modified according to the size of the problem task. The size of the part of the visual scene for representing objects was set to be twice the size of the maximum number of objects being counted, allowing for spaces to be included between neighbouring objects. (Ahmad et al., 2002, pp. 190–191)

What remains unclear is how the 72 locations in the scale invariant module were mapped onto the 44 units in the reduced-dimensionality representation. To add to

the confusion, the only example shown in the paper (Ahmad et al., 2002, Table 8) suggests that the 44 units may have been allocated from left to right, resulting in a thermometer-like representation, albeit with 1-unit spaces between the activated units. In contrast, in the earlier paper describing the earlier incarnation of the model (Ahmad & Bale, 2001, Figure 4), the objects in the provided example are clearly allowed to be located anywhere in the vector, however without the imposed spaces. In any case, the visual scene representation passed on to the counting module is a 44-dimensional vector in which the activated units correspond (in the way that is unclear) to the objects arranged in a single row.

In addition to the 44-dimensional representation of the visual scene, the counting module employs two additional representations. First, the action of pointing to the next object is encoded using a 45-element vector. The first 44 elements of this vector correspond to the 44 locations in the visual scene vector with the activated unit representing pointing to the associated location. The additional 45th unit represents a special ‘no pointing’ action meant to be executed when the counting is finished. Second, an 18-element vector encodes the number words using a simplified phonetic coding in which English number words are divided into common parts and activated units in the vector indicate the presence of the corresponding word segments.

The counting module of the model of Ahmad et al. employs a mixture-of-experts architecture with recurrence at two levels. There are two underlying experts, one feed-forward and one recurrent (at the local level). Both use the same format of input and output. At the input the experts accept a concatenation of the visual and pointing representations. At the output the experts provide a concatenation of the current number word and of the next pointing act to be performed. A system of gating networks decides which of the two experts is allowed to populate which parts of the final output vector. The part of the output vector representing pointing is fed back to the input of the counting module in the next time step, providing recurrence at the global level. Finally, a subset of the composite output vector

elements, namely the number word part and the ‘no pointing’ unit, constitutes the output of the counting module which is passed on to the output module.

The first of the two experts employed in the mixture-of-experts system of the counting module is a feed-forward 3-layer network with 20 units in the hidden layer. The second one is a recurrent network of the Jordan type (Jordan, 1986) with 9 units in the hidden layer and 18 state units without self-recurrent connections, representing the number words as described earlier. The inclusion of two different experts in the architecture was motivated by the fact that in the presence of the global recurrent feedback of the pointing action, a feed-forward mechanism should suffice to determine the next object to be pointed to, whereas generating subsequent number words would still require a mechanism recurrent at a lower level. The gating mechanism is then expected to allocate each of these two sub-tasks to the most suitable expert.

The task of the final module of the model, the output module, is to provide the ultimate task response, that is to produce the number word representing the result of counting, once this process is finished. This module is implemented as a single-layer Madaline network (Widrow & Lehr, 1990) with 19 inputs and 64 outputs. The input layer accepts the output of the counting module (18 units representing a number word plus the ‘no pointing’ unit). The output layer implements the topological mapping of the phonetically-encoded number words, compatible with the one used by the subitising model described in the same paper (see section 3.2.3).

The various elements of the model of Ahmad et al. were trained individually, using dedicated techniques. For the elements with recurrent connections the teacher forcing method was used, in which the correct values of the state units in each time step are provided instead of feeding back the actual output of the network from the previous time step. It is not clear whether the task decomposition in the mixture-of-experts architecture was incorporated in the training regime (and thus affected the training data available to each expert), or was the gating network trained after the fully developed experts have already been in place. The output module was trained

to forward the number word as the model response only when the ‘no pointing’ input has been activated. The latter acted therefore as an explicit indication of the end of counting.

Ahmad et al. performed two kinds of simulations of their counting model. They compared its performance with children’s counting accuracy in terms of the production of the number words and in terms of pointing. Their analysis included the assessment of the model’s ‘developmental trajectory’, implemented by evaluating the model’s behaviour after exposing it to a variable amount of training.

In assessing the model’s counting list, Ahmad et al. used the behavioural data found in Fuson et al. (1982). According to Fuson et al., the children’s counting list can be divided into three parts: stable conventional (i.e. correct and consistent), stable non-conventional (incorrect but consistent), and unstable (incorrect and inconsistent). Ahmad et al. state that in the counting list produced by their model all three subsequences could have been distinguished. The example provided in the paper (Ahmad et al., 2002, table 9) shows however only that the counting list could be divided into a stable conventional and unstable part, with the former becoming longer in the course of the model development. It shows no evidence that the stable non-conventional part was present. Ahmad et al. report also that another characteristic element of the children’s learning to count, namely the difficulty with learning irregular number words, was exhibited by their model. Specific data in support of this result were however not provided.

The second aspect of the model’s performance tested in the simulations was the types of the pointing errors and the relative frequency of their occurrence. To that end, the behaviour of the model was compared with that of children as reported by Fuson (1988). Ahmad et al. distinguished the following types of the pointing errors in their model’s behaviour: *object skipped*, *multiple count*, *no object* (when the model pointed to an empty location), and *stopped early*. The frequency of the errors made by the model over the course of training was fit to that of the children based on the best-matching error type, the *object skipped* error. The comparison (Ahmad et al.,

2002, figure 5) showed that the number of errors made by the model decreased along with the progress of training. The relative frequencies of the committed errors were however not consistent with the human data. Most importantly, an error that is rarely committed by the children (the *no object* error) has been the most frequently made error by the model.

The model of Ahmad et al. (2002) is the most ambitious and exhaustive existing effort to simulate the children's ability to count. It incorporates all three key elements of the counting process: vision, pointing and number words production. There are however areas in which improvements to the approach taken by Ahmad et al. can be made. I discuss these in detail when proposing my own model of counting in chapter 5.

3.4 Models of the SNARC Effect

The first model of the SNARC effect was proposed by Gevers, Verguts et al. (2006). It was based on the previous work by the same authors discussed earlier (Verguts et al., 2005, see section 3.2.5). The overall design principle as well as the main elements of the model remained unchanged, but in order to account for the SNARC effect, a few additions to the model were made. First, a *response layer* was added after the existing task-dependent decision layers. This new layer consisted of two units with lateral inhibition, which represented the left- and right-hand response, respectively. This enabled mapping of the task response to the either response hand (e.g. press left when even vs. press left when odd) which is the key element of the experimental design aimed at demonstrating the SNARC effect. Second, it was assumed that in all simulated tasks (number comparison, parity judgement and the arbitrary mapping task), in addition to the task-dependent decision pathway there is a second, *automatic pathway*, in which, independently of the task, every presented number is classified either as small or large, what in turn primes the response with the left or right hand. This was realised by assuming that in tasks

which require only one number input (i.e. the parity judgement and the arbitrary mapping task), the second number is implicitly present in form of the mean of the interval of numbers presented in the task and acts as the small/large comparison standard. This solution is plausible because firstly, the subjects in the behavioural experiments are usually informed about the range of numbers they are going to be presented with and secondly, because the task context (i.e. the range of considered numbers) is known to affect the shape of the SNARC effect response profile.

This dual-route architecture of the model of Gevers, Verguts et al. (2006) accounted for the SNARC effect in the following way. When the task response was congruent with the automatic small-left/large-right mapping, the response time was shortened as the correct hand has been additionally primed via the automatic pathway. Conversely, when the task response was incongruent with the automatic mapping, the response time was prolonged because the response mapping had to ‘work against’ the activations present in the automatic pathway. As the result, the responses to small numbers with the right hand were slower than those with the left hand, producing the SNARC effect.

The weights in the model were partially trained and partially set by hand. The values of the trained weights were taken directly from the earlier simulations (Verguts et al., 2005). The remaining weights, which included the weights in the automatic pathway, were adjusted so that the desired magnitude of the SNARC effect was obtained.

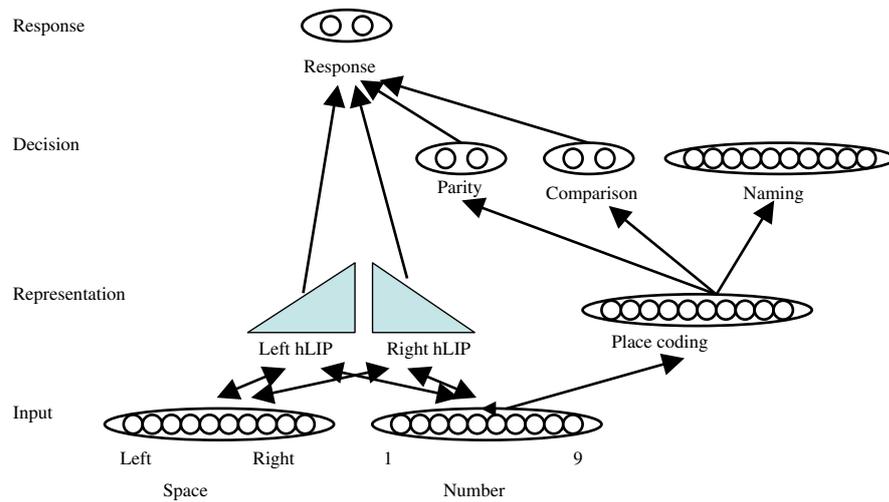
In simulations Gevers, Verguts et al. (2006) demonstrated that their model exhibits the SNARC effect in the number comparison and parity judgements tasks. In addition to providing the desired pattern of the response times, the model reproduced another feature of the effect, namely the increase of the magnitude of the effect along with the response time (what was modelled by scaling the magnitudes of the activations in the input layers). The main finding however was that the shape of the SNARC effect response profile was different in the two tasks — continuous over the considered number range for the parity judgement, and categorical (with

an evident step around the comparison standard) for the number comparison. This pattern was then confirmed to be present in human response times in a behavioural study, the results of which were reported in the same paper. The model explained the categorical shape of the response time profile in the comparison task as the interaction between the SNARC effect and the numerical distance effect — since for closer numbers the response time is longer as the result of the distance effect and with the longer response time the magnitude of the SNARC effect increases, numbers close to the comparison standard take disproportionately longer to process, resulting in the step-shaped response time profile. Since the model further predicted that a continuous SNARC effect should be found in tasks with an arbitrary magnitude-to-response mapping, this was investigated in another behavioural experiment, with a positive result.

The model of Gevers, Verguts et al. (2006) was further extended by Chen and Verguts (2010). Here, in addition to the pathway which in the earlier model automatically classified a number as small or large, an explicit representation of space and of the association of numbers with space were introduced (see figure 2). The representation of space was implemented using three clusters of units. The first one, the *space layer*, encoded the location of stimuli in the eye-centred frame of reference with each of the 10 units in the layer corresponding to one spatial location, from left to right. This layer acted as an input to the model, e.g. indicating the spatial location where a visual target is presented on a screen, or representing a horizontal line in the line bisection task. The remaining two clusters, the *left* hLIP and the *right* hLIP, were introduced as corresponding to the hypothetical human homologue of the lateral intra-parietal area in primates. The biological motivations for introducing these two layers can be found in the paper (Chen & Verguts, 2010, p. 220); their most important characteristic is that the left and right hLIP over-represent the right and left side of space, respectively. More specifically, out of the 45 units in the left hLIP layer, 9 have the peak response for the rightmost unit of the space layer, 8 to the one-but-rightmost unit, and so forth up to the the 1 unit having the peak

response for the leftmost unit of the space layer. For the right hLIP this neuronal gradient is reversed. The association of numbers and space was realised in the model as connections between the number layer and the left and right hLIP. These links had an analogous structure to the connections between the space layer and hLIPs described above, with small numbers connecting more strongly to the right hLIP and large numbers to the left hLIP. Such a pattern of connections between numbers and space was assumed to arise as a result of the cultural determinants but in the actual model was implemented by hand (Chen & Verguts, 2010, pp. 220, 237). Finally, the left and right hLIPs were connected to the right and left unit in the response layer respectively, thus providing the secondary source of the SNARC effect in the model: for instance the presentation of a small number activated the right hLIP more strongly than the left one, as a consequence priming the left-handed response.

Having the space input layer as well as the left and right hLIP modules in the model made it possible to extend the range of the behavioural effects that the model was able to exhibit. In particular, Chen and Verguts (2010) demonstrated that their model, in addition to reproducing the most important effects accounted for by the previous model (namely the number size and numerical distance effects in the number comparison task, the symmetrical priming patterns in the number naming and parity judgement tasks and the SNARC effect in the parity judgement and number comparison tasks), allowed to simulate the Posner-SNARC effect, the patterns of number comparison response times as well as the SNARC effect in patients with left neglect, and, finally, both associations and dissociations in the physical and mental number line bisection tasks (see section 2.2.3). The response times of the model were simulated explicitly, but were read out from different modules, depending on the task. Simulating the line bisection tasks required new response layers to be trained and added to the model. The neglect was modelled by removing or selectively damaging certain parts of the model (e.g. left neglect involved removing the right hLIP). Overall, the model's behaviour fit an impressive corpus of experimental data to a satisfactory degree, thus establishing the state-of-the-art in the modelling of



(a) Structure of the complete number-space model. The lower two layers are input layers; they map to space (hLIP) and number (place coding) representations at the second level. The next level contains decision units for three different tasks (parity judgement, number comparison, and number naming). At the decision level, the number of units in a given layer corresponds to the number of possible responses for each task (odd and even in parity judgement; small and large in number comparison; verbal responses ‘one’ to ‘nine’ in number naming). At the response level, there are two units, corresponding to a left and right response, respectively. For the number naming task, the decision and response levels coincide.

Figure 2: Architecture of the model of spatial-numerical associations by Chen and Verguts (2010). Reprinted from *Cognitive Psychology*, 60(3), Chen, Q. & Verguts, T., *Beyond the mental number line: A neural network model of number-space interactions*, p. 222, Copyright 2010, with permission from Elsevier. The original caption is quoted in (a).

the interactions between numbers and space.

3.5 Research on Gestures in Humanoid Robots

Gestures in general are of great importance in the context of humanoid robotics, especially from the point of view of human-robot interaction (Breazeal, 2002). On one hand, the idea to build robots whose physical appearance resembles humans is intended to facilitate the interaction with robots by catering on our strong tendency toward anthropomorphisation (Duffy, 2003). On the other hand, body language, and gestures made with hands in particular, are of paramount importance for human communication (Kendon, 1980; McNeill, 1992; Goldin-Meadow, 2003; Hostetter, 2011). It is not surprising therefore that the amount of literature on gestures in human-robot interaction is very large. Considering that the communicative gestures are not the focus of the present work, a thorough review of this broad field is far beyond the scope of this thesis. Nevertheless, recognising the relevance of the topic, a few selected prominent robotic studies that looked at gestures are presented below.

Naturally, much of the early work on gestures in humanoid robots focused on tackling the engineering issues that arise already at the level of gesture production. As an example one can take the work of Marjanović, Scassellati and Williamson (1996) who implemented a visually-guided pointing system in the humanoid robot Cog. The task of the system was to saccade the robot's eyes to the target, and then generate a smooth trajectory for the robot's arm that would bring it into the configuration corresponding to pointing to the target. This involved autonomous learning of the mapping between the space of the plane of the image perceived by the robot's wide-angle camera and the robot-centred frame of reference in the space of the pan and tilt camera movements in order to achieve the saccadic behaviour. The realisation of the robot arm movements was based on linear interpolation between four available 'postural primitives' and a bi-directional mapping between the visual and motor spaces. Both the 'ballistic map' (the mapping from the eye position to

the arm position) and the ‘forward map’ (the mapping from the arm position to the eye position) were implemented using the radial basis function approach, and were trained simultaneously with the least-mean-squares gradient descent method based on sample reaching movements and the robot’s visual observation of its own arm. The main limitations of the system were that it did not use all of the degrees of freedom available in the robot and that the workspace covered by the pointing was only two-dimensional.

An important category of communicative gestures are the *representational gestures*, which include

movements that represent the content of speech by pointing to a referent in the physical environment (deictic gestures), depicting a referent with the motion or shape of the hands (iconic gestures), or depicting a concrete referent or indicating a spatial location for an abstract idea (metaphoric gestures) (Hostetter & Alibali, 2008, p. 495).

Since such gestures co-occur with and semantically supplement speech, the two together constitute a form of multi-modal communication. Salem, Kopp, Wachsmuth, Rohlfing and Joublin (2012) proposed a system for the production of representational gestures accompanying synthesised speech on a humanoid robot ASIMO (Sakagami et al., 2002) aimed at achieving more natural human-robot interaction (see also Salem, 2012). The system was based on segmentation of the continuous multimodal communication signal into chunks representing single ideas. Temporal synchronisation within chunks was achieved by adapting the gesture production to the timing dictated by the structure of the synthesised speech. The repertoire of the available gestures included iconic gestures (which illustrated the shapes or sizes of objects), pantomimic gestures (which demonstrated the activity the robot was referring to verbally), as well as deictic gestures (location indications). The effectiveness of the system was assessed empirically in a study with human participants who interacted with the robot in three conditions (speech without gestures, gestures congruent with speech and gestures incongruent with speech) and evaluated the experience afterwards. Multimodal communication turned out to be associated with higher

ratings of the quality of the interaction with the robot, and, interestingly, on most of the criteria (6 out of 8) the scores were higher in the incongruent than in the congruent speech-gesture condition. This led the authors to conclude that ‘imperfect’ communicative behaviour of a robot may lead to even stronger positive response in humans.

In order to facilitate natural, bi-directional human-robot interaction, in addition to being able to produce gestures, the robots need the ability to interpret the body language of others. Hafner and Kaplan (2005) considered the case of the pointing gestures as a means of non-verbally directing the other party’s attention. Possession of such a skill is important for example in order to bootstrap influencing the attention verbally. Hafner and Kaplan considered a scenario with two Sony AIBO robots. The first robot, the ‘adult’, performed pointing gestures in an attempt to direct the attention of the other robot, the ‘child’, to an object located either to the left or to the right. The task of the child robot was to look at the adult robot and learn to exploit its indications in order to find the object being pointed to. Gesture recognition was achieved using a multi-layer perceptron classifier, trained on a set of 2300 sample images taken with the camera of the child robot while it looked at the parent robot. The images were pre-processed by extracting the features that have been determined to be the most suitable for the classification using pruning methods. The classifier was then trained via backpropagation. The achieved success rate of discriminating pointing to the right from pointing to the left reached 95.96% using the selected subset of the image features, and 98.83% when all considered features were used. While limited in several ways, the study was, according to its authors, the first attempt to teach a robot to interpret the pointing gestures of another robot.

The progress in computer vision algorithms as well as in sensing technology enables the construction of increasingly more sophisticated systems of gesture recognition. Recently, Fanello, Gori, Metta and Odone (2013) proposed a system for recognition of 3-dimensional actions in the context of human-machine interaction,

capable of one-shot learning (i.e. learning based on a single example) and classifying the actions in real-time. In addition to colour video data, the system relies on the depth information (which can be obtained e.g. using the Microsoft™ Kinect™ sensor, *Kinect for Windows website*, 2013) to identify the region of interest in the input image. According to the authors this significantly reduces the complexity of the system architecture and permits the use of simpler features to describe the performed action without sacrificing the discriminative power. The features in terms of which the actions in the region of interest are described are the 3D histogram of flow and the global histogram of oriented gradient. These are followed by a sparse-coding stage which finally feeds the data to the linear Support Vector Machine-based classifier. Optionally, the system can also use the body part tracking data (available as well through the Microsoft™ Kinect™ sensor), in order to isolate the hands of the person and thus allow the recognition of hand gestures. The system of Fanello et al. is capable of high classification accuracy while retaining real-time performance (25 frames per second on a 2.4 GHz personal computer). It has been applied for example in a human-robot interaction scenario in which a human competes with the iCub humanoid robot (see section 4.3) in a gesture memorisation game (Gori, Fanello, Metta & Odone, 2012). In the game the players take turns in performing a sequence of gestures, always starting with repeating the gestures made by the opponent in the previous turn and including an additional gesture at the end. The challenge for the human is to memorise the prolonging sequence of gestures correctly, while the robot may lose due to the failures in the recognition process. The game has been awarded the second place in the ChaLearn 2011/2012 One-shot-learning Gesture Challenge demo competition (*ChaLearn Gesture Challenge website*, 2013).

3.6 Computational Modelling in Mathematical Cognition — Summary

In the present chapter the past endeavours to computationally model different facets of human mathematical cognition have been reviewed. These studies provide the context for my own modelling work. The models recalled in this chapter are going to be referred to often when I describe my models. Many design decisions and implementational solutions in the present work were based on, or took inspiration from, the experience and results obtained by other researchers.

At this point it is appropriate to take a look at the past modelling research in mathematical cognition from a broader perspective and try to analyse it from the point of view of the main goal of this thesis, expressed through the four research questions posed in the introduction. This will help me to show where in the cognitive modelling research landscape are my models situated and to emphasise the original contribution of the present work. From the point of view of modelling learning to count in humanoid robots, it is important to point out the following about the past modelling research in numerical cognition:

- none of the discussed models took advantage of the cognitive robotics methods. All reviewed models are purely computational, and therefore ‘disembodied’². Contrasting this fact with the ample evidence that human mathematical thinking has embodied foundations presented in section 2.2, reveals an important and promising open avenue that can lead to improving the state of the art in the field. The present work is the first effort to fill in this niche;
- considerable body of work exists that focused on modelling of the mental representation of magnitude (reviewed in section 3.2), so it is more than appropriate for any future work, including the one presented herein, to refer to and build on top of this past experience, avoiding thus ‘reinventing the wheel’;

² Chen and Verguts (2010) included ‘embodied’ features in their model, although in a rather abstract way, in form of the lateralised representation of space.

- as shown in section 3.3, not many models exist that have looked at the counting skill, and most of them focused on its very narrow aspects. For example, the model of Ma and Hirai (1989) investigated the behavioural effects connected with the learning of the count list, but did not consider how it may affect the process of learning to count as a whole, and therefore did not address the research question 1. The most complete model of counting proposed to date (that of Ahmad et al., 2002) is complex and unfortunately its description is, in certain aspects, unclear. Although it incorporated a representation of gestures, and thus was, in principle, suitable for addressing questions similar to the research questions 2 and 3, this has not been made. In my own model I propose several alternative solutions to those employed by Ahmad et al., especially concerning the representation of the gestures;
- Verguts et al. published two models of the SNARC effect (described in section 3.4), which explained an impressive amount of experimental data. However, since the weights of some of the crucial components of these models were set by hand, they are, in their original formulation, not suitable for the investigation of the question about the ontogeny of the spatial-numerical associations (the research question 4). By extending these models with a more realistic representation of embodiment and by proposing an associated development process, I demonstrate that the spatial-numerical associations may appear as the result of the spatial biases present when children learn to count, consistent with a hypothesis which has appeared in the experimental literature in parallel with my work (Opfer & Furlong, 2011).

This and the previous chapter provided the appropriate background in the areas of psychology and cognitive modelling. The next chapter sets the stage for the description of my models by introducing the methods employed in the present work.

Chapter 4

Methods

This chapter contains an overview of the methodological aspects of the research presented in this thesis. First, I outline the embodied view of cognition by reviewing a number of more specific claims that can be identified under this general term. This is followed by a short introduction to developmental cognitive robotics, a paradigm that, in line with the philosophy of embodiment, supplements computational cognitive modelling with an artificial, robotic body. Next, the particular robotic platform that has been used throughout the modelling experiments described in this thesis is reviewed, namely the iCub humanoid robot. Subsequently, the artificial neural network frameworks I employed in formulating my models are presented. Finally, the chapter concludes with an example of an experimental set-up used in a behavioural study aimed at investigating the role of gestures in learning to count in children, which has inspired the design of some of the experiments conducted in this thesis.

4.1 Embodied Cognition

Embodied cognition is a line of thought in cognitive science according to which, in the general sense, the mind cannot be understood without taking into account the body and its interactions with the outside world (A. Clark, 1998). One of the main arguments that can be put forward to support such a view is that the human brain shares a lot in common with the brains of lower animals, which are, for the most part,

devoted to sensorimotor processing. Since the cognitive functions of animals consist mostly of on-line interactions with their environment, it is sensible to assume that human cognition has its roots in such perceptual and motor processing and builds on top of it (M. Wilson, 2002). Although this idea is certainly not new, it has been gaining popularity especially during the past two decades.

M. Wilson (2002, p. 626) summarises the six most prominent views that can be distinguished under the broad term of embodied cognition as follows:

Cognition is situated, which means that the cognitive processes are accompanied by a constant stream of incoming task-relevant perceptual information and, at the same time, the executed actions affect the surrounding environment (Steels & Brooks, 1995). In other words, a constant interaction takes place between the cognitive agent and the things that are the subject of the cognitive activity;

Cognition is time pressured. The fact that cognitive agents live in a constantly-changing environment and have to deal with dynamic situations that happen in ‘real time’ imposes strong constraints on the computational aspects of the cognitive processes (Pfeifer & Scheier, 1999). It is sometimes stated that the time pressures create the ‘representational bottleneck’ in the sense that due to the little amount of time available, building a complete internal model of the environment and using it to construct a plan of action is not a feasible solution;

We off-load cognitive work onto the environment. In order to overcome the representational bottleneck mentioned above, instead of encoding and retaining information relevant to the situation in the working memory, we use our environment in strategic ways that allow us to reduce the cognitive workload (Kirsh & Maglio, 1994). Brooks (1991, p. 139) referred to this famously as the world being ‘its own best model’;

The environment is part of the cognitive system, which in its strong form

means it is not sufficient to approach cognition solely as an activity of the mind, but the entire cognitive situation, including the cognitive agent and its environment, must be studied as one system (A. Clark, 1998). According to M. Wilson (2002, pp. 629–631) this is the most controversial of the six claims;

Cognition is for action. In other words, one has to approach the study of cognitive mechanisms, such as perception or memory, focusing on the ways they help to achieve adaptive behaviour (Glenberg, 1997). This is connected with the assumption that ‘we conceptualize objects and situations in terms of their functional relevance to us, rather than neutrally or “as they really are”’ (M. Wilson, 2002, p. 631);

Off-line cognition is body based. This states that the cognitive activities that have been often regarded as ‘abstract’ may actually be supported by the sensorimotor functions executed in a covert way. In order to achieve this, it must be possible to decouple somehow the capacities primarily devoted to dealing with perception and action from their ‘original’ inputs and outputs, so that they may assist the process of thought (Glenberg, 1997). Examples of cognitive processes that are likely based on such sensorimotor simulations include mental imagery, working, episodic, and implicit memory, as well as reasoning and problem-solving (M. Wilson, 2002, pp. 633–634);

Although these claims vary with respect to the strength and the degree of controversy they rise, such a decomposition is useful to underscore why is it appropriate to look at learning to count — around which the four research questions addressed in this thesis revolve — from the point of view of embodied cognition.

The ample evidence that human mathematical thinking in general, and counting in particular, is a prime example of embodied cognition, has been reviewed in chapter 2. On the most general level, the fact that the rules of arithmetic may be formulated as a conceptual metaphor of simple interactions with physical entities, such as collections or measuring units, is perfectly in line with the claim about

the situatedness of cognition. All four conceptual metaphors proposed by Lakoff and Núñez (2000) refer to scenarios and activities that are encountered countless times throughout our early development (such as playing with a collection of toys) or are indeed fundamental to our existence (as is motor control). The theoretical considerations of Lakoff and Núñez, supplemented by many lines of evidence for the link between ‘concrete’ spatial and ‘abstract’ numerical representations in the brain suggest the involvement of body-based representations as a substrate for forming abstract mathematical concepts. This is consistent with the statement that off-line (in other words, abstract) cognition is body based. Perhaps the most striking evidence in support of the latter claim is the study of Andres et al. (2007), where covert involvement of finger motor circuits in sequential enumeration tasks was demonstrated. The claim about off-loading cognitive work onto the environment is directly relevant to the first group of the hypotheses about the contribution of gestures to learning to count reviewed in chapter 2, which focus on overcoming limitations in available cognitive resources (see section 2.2.2). Finally, the second group of the hypotheses, which emphasised the coordinative role of the gestures and the hypothetical transfer of a motor competence to a conceptual one, resonates with the claims about the relations between cognition and action. There is little doubt therefore that embodied cognition provides an appropriate theoretical framework within which learning to count and phenomena associated with it should be considered.

4.2 Developmental Cognitive Robotics

The concepts connected with embodied cognition are among the essential themes in *developmental cognitive robotics*. Developmental cognitive robotics has been introduced by Asada et al. (2001) as a novel humanoid robots design principle. Its primary aim is to ‘understand the cognitive developmental processes that an intelligent robot would require and how to realize them in a physical entity’ (Asada et

al., 2001, p. 185). As a methodology, developmental cognitive robotics opposes the traditional ‘engineering’ approaches to build robots, which involve explicit programming of the robot functionality and its control algorithms by the designer, based on his knowledge of the physical characteristics of the robot and of the all possible situations in which it is intended to operate. Instead, Asada et al. advocate that the robot’s control structures should ‘reflect the robot’s own process of understanding through interactions with the environment’ (Asada et al., 2001, p. 185).

Two fundamental problems that are addressed in the developmental cognitive robotics design principle are: how to design the robot’s ‘brain’, or control structure, that would be capable of learning and developing, and what kind of a social learning environment should be established in order to support the development of the cognitive processes of interest. While a lot of progress with respect to the former issue has been achieved through the years of research in robotics and artificial intelligence, Asada et al. (2001) put a lot of emphasis on the role of the social set-up in human development. This is motivated by the fact that the tutoring delivered to children by their parents, teachers, etc. is characterised by a great degree of adaptivity. In other words, it is adjusted adequately to the child’s current level of competence, which may be the key to assuring the optimal learning progress. Accordingly, the appropriate set-up of the learning environment, that allows the robot to gradually learn more and more complex tasks and handle increasingly unpredictable situations, is a prominent part of the developmental cognitive robotics methodology.

Developmental cognitive robotics has strong interdisciplinary links with developmental psychology, neuroscience, and cognitive science (Asada et al., 2001; Cangelosi & Schlesinger, to appear). It can be argued that it holds a promise of twofold benefits that it can bring to the scientific society (Lungarella, Metta, Pfeifer & Sandini, 2003). On the one hand, being a robot design methodology, it is hoped to help to overcome the challenges that the roboticists have been facing in creating intelligent machines. On the other hand, developmental cognitive robotics can be viewed as a framework for implementing models of human cognition. As it supports the use of

computational approaches to modelling with an artificial body, it permits a fuller exploration of the claims put forward by the embodied view of cognition.

Although a relatively young discipline, developmental cognitive robotics is dynamically expanding. While at first most of the research in the field focused on the modelling of the development of motor control (Lungarella et al., 2003), more and more studies appear that tackle the issues of higher-level cognition (Cangelosi & Schlesinger, to appear).

In the context of the present work, developmental cognitive robotics is a crucial part of the employed experimental methodology. As reflected by the research questions posed at the beginning of this thesis, the phenomena considered herein — the contribution of the counting gestures to learning to count and the spatial-numerical associations — have a pronounced embodied character. Therefore, the complementary role developmental cognitive robotics plays to the computational modelling of cognition will be, from the point of view of present research, a significant and important extension.

4.3 iCub — Humanoid Robotic Platform

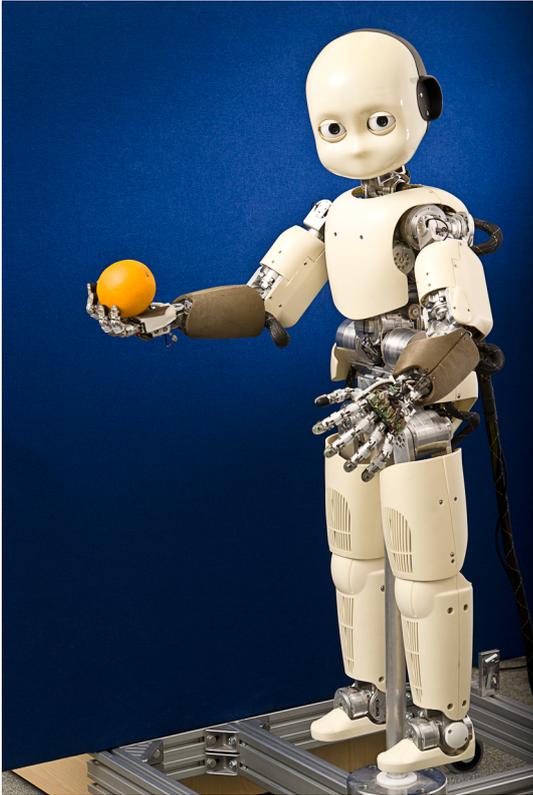
As outlined above, an important aspect of the developmental cognitive robotics paradigm is the inclusion of an artificial body in the cognitive modelling process. Consequently, designing a neuro-robotics experiment requires settling on the choice of a particular robot that will be appropriate for the research. The robotic platform used throughout the experiments described herein is the *iCub humanoid robot* (Metta, Sandini, Vernon, Natale & Nori, 2008; Metta et al., 2010, see figure 3).

The iCub platform, including the complete robot itself, as well as the accompanying software, has been developed from scratch (Metta et al., 2010) within the European Commission-funded, sixth framework programme project RobotCub, headed by the Italian Institute of Technology, and completed in 2010 (*RobotCub project website*, 2013). At the time of writing, various versions of iCub are being used

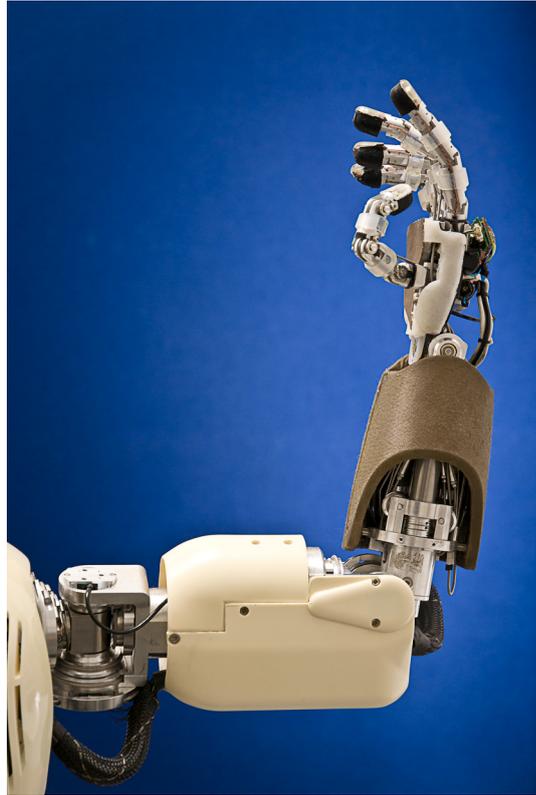
in more than 20 research laboratories worldwide (*iCub website*, 2013), and ongoing projects aim at further developing the platform by providing it with additional capabilities, like whole-body motion interaction with multiple contacts (*CoDyCo project website*, 2013).

One of the primary motivations that guided the design of the iCub was the belief that ‘the physical instantiation of [...] [the models of the development of human cognition] in a behaving humanoid robot’ is one of the crucial ingredients necessary to allow the construction of cognitive systems to progress (Metta et al., 2010, p. 1126). The RobotCub project aimed therefore at providing a platform that would fulfil the present and future requirements of the researchers working on the development of cognition (Vernon, von Hofsten & Fadiga, 2010). In line with the view that object grasping is central to human cognition (Arbib, 2002b), strong emphasis has been put on the capabilities of the robot connected with manipulation.

On the mechatronics side, the iCub robot, shown in figure 3a, is approximately 1 meter tall and weights around 22 kilogrammes, what is intended to resemble the body dimensions of a 3- to 4-years-old child (Metta et al., 2010). Its upper body incorporates the total of 38 actuated degrees of freedom (DOF). 6 of these are allocated to the head, 3 implementing the full range of neck motions, and 3 supporting the movements of the eyes, including vergence. Each of the arms of the robot has 7 DOF: 3 in the shoulder joint, and 2 in both elbow and wrist. The robot’s hands have 9 DOF, allowing for considerable dexterity and reflecting the design emphasis on manipulation (see figure 3b). The first three fingers can be controlled independently, while the ‘ring’ and ‘pinkie’ fingers are linked together under 1 DOF, providing additional support and stability for grasping. The manipulation capabilities are further extended by 3 DOF in the robot waist that increase the reachable working space. The original version of the iCub was designed to support crawling and sitting rather than walking, and therefore each of the legs has 6 DOF. This brings the grand total of the actuated DOF of the robot to 53. The legs of the robot were not used in the present study.



(a)



(b)



(c)

Figure 3: iCub, the humanoid robot used in the present study. (a) the robot at the disposal of Plymouth University that has been used throughout the experiments described herein; (b) close-up of the iCub's 16-DOF arm; (c) the virtual counterpart of the robot as seen in the simulation software.

For the ‘static’ robotic experiments that do not involve crawling or sitting, the robot is used attached to a stand by its hip, as depicted in figure 3a. As iCub was not meant to be fully autonomous in terms of power supply and computational power (Metta et al., 2010), the connection with the external facilities is achieved through the ‘umbilical cord’ attached to the back of the robot. As mentioned earlier, a project that aims at upgrading the iCub’s lower body in order to enable using it in a free-standing set-up is under way (Eljaik et al., 2013; *CoDyCo project website*, 2013). For the time being, for experiments requiring mobility an optional mobile base for the robot, called *iKart*, is available.

The interaction of the iCub with its environment is realised through a range of sensors available on-board. Vision and audition are provided by the digital cameras located in the robot’s eyes and the microphones on both sides of its head, respectively. Proprioceptive sensing includes the inertial sensors located in the head (3 gyroscopes, 3 linear accelerometers and a compass), as well as the force/torque and position sensors in the body joints. Recently, the repertoire of the iCub’s sensing capabilities has been extended with an artificial modular skin that can be mounted on the robot’s fingers, palms, forearms and torso (Schmitz et al., 2011).

The entire design of the robot is open-source, which includes the robot’s software (Metta et al., 2010). The iCub-specific software architecture is developed on top of a more general middleware called YARP (Yet Another Robot Platform, Metta, Fitzpatrick & Natale, 2006). The latter provides operating system-independent facilities for interfacing with the robot hardware as well as for implementing the robot control software in a modular way. The range of software available for the iCub robot is being constantly extended by an active community of users and researchers (*iCub website*, 2013). Among those, certain software tools were of particular importance for the present study. Firstly, the *iCub simulator* (Tikhanoff et al., 2008; Tikhanoff, Cangelosi & Metta, 2011) provides a virtual copy of the robot than can be used for developing and testing of the control software in a rapid and safe way without the need of the physical access to the robot (see figure 3c). The precision of the

simulation is usually sufficient to allow for an easy transition of the ready software to the real robot. Secondly, the YARP *Cartesian interface* (Pattacini, Nori, Natale, Metta & Sandini, 2010), together with its iCub-specific implementation, provides a solution for the robot’s inverse kinematics that enables the user to control the robot directly in its working space, using the Cartesian coordinates.

Overall, the motivations for using the iCub as the principal robot platform for the present study can be summarised as follows:

- large number of the degrees of freedom in the upper body of the robot together with the humanoid design allow for faithful modelling of the counting gestures;
- required capabilities (robot simulation, Cartesian control of the robot’s hands and gaze, position control and sensing) available out-of-the-box;
- open-source robot software can be extended as needed;
- easy and flexible interfacing of the available robot software with custom control software.

4.4 Artificial Neural Network Models Employed in the Study

This section is an overview of the specialised artificial neural network frameworks that are employed throughout the present study. It is assumed that the reader is familiar with the fundamental concepts connected with neural modelling, such as the activation function, multi-layer perceptron or backpropagation. A good introduction to these concepts can be found in the classical work of Rumelhart and McClelland (1986) or in a more recent book by Levine (2000).

4.4.1 Simple Recurrent Artificial Neural Networks

Discrete-time feed-forward neural networks are static, in the sense that, for a given input vector, a well-defined corresponding output vector is computed that does not

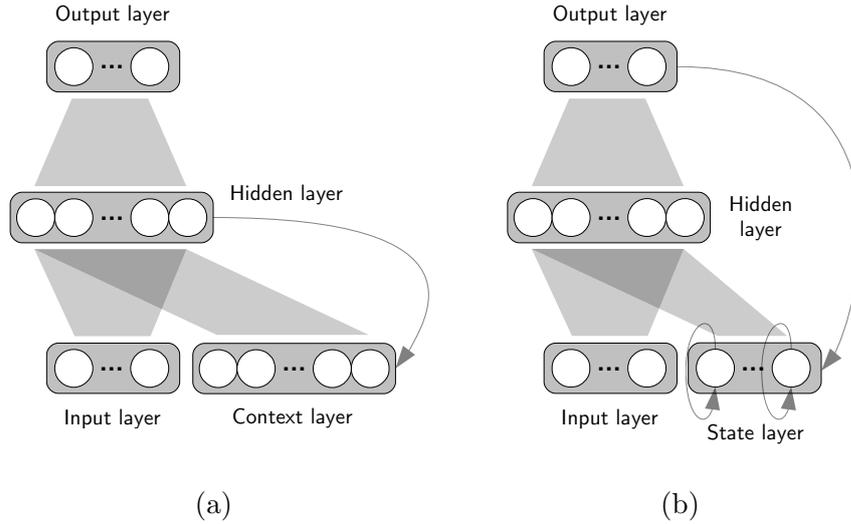


Figure 4: Elman and Jordan networks. Grey areas represent all-to-all feed-forward connections, arrows represent recurrent connections. In the Elman network (a), memory is realised by retaining a copy of the activations of the hidden units from the previous time step. In the Jordan network (b), memory is provided by the state units which compute a weighted average of the output of the network over time.

change over time, nor depends on the history of the previous inputs provided to the network. However, in the context of simulating biological systems or designing the control architectures for robots, the capability to process signals that unfold in time is desired. One of the ways in which a discrete-time feed-forward neural network can be extended to achieve such a capability, is providing it with a form of memory. Memory can be introduced to a discrete-time neural network using recurrent connections, that is connections that create cycles in the network connectivity graph. Figuratively, such connections ‘look back in time’, in other words they refer to the activation values of their input units determined in the previous time step of the simulation. Two classical architectures that make use of this concept are the *Elman network* (Elman, 1990) and the *Jordan network* (Jordan, 1986).

The Elman network (figure 4a) is an extension of a standard three-layer feed-forward network. In order to provide the network with memory, a layer of units, called the *context units*, that connect to the hidden layer, is added to the network. In the first time step of the network simulation, these units have a fixed activation value of 0.5. In every subsequent time step, they contain the activation values of

the units of the hidden layer from the previous time step. This corresponds to introducing fixed (not modifiable through training) recurrent connections between the hidden layer and the context layer, the weights of which form an identity matrix. This, in turn, is equivalent to introducing modifiable recurrent connections from the hidden layer to itself with a fully-connected topology. Since the output of the Elman network is a function of its inputs and its internal state from the previous time step, such a network has an inherent capability to process temporal signals. The Elman network has been applied primarily in the context of language processing, but, as reviewed in chapter 3, studies that employed this architecture in the context of counting exist as well (Hoekstra, 1992; Rodriguez et al., 1999).

The architecture of the Jordan network (figure 4b) is somewhat more elaborate than that of Elman. The assumption here is that the network generates a sequence of *actions* (output vectors) based on a *plan* provided as a fixed input to the network. The three-layer feed-forward network architecture is extended with a *state layer*. The state layer is connected to the hidden layer and accepts recurrent connections from the network output and from itself. Similarly to the Elman network, these recurrent connections are not modifiable through training. In the classical formulation, the state units implement an exponentially weighted average of the past outputs of the network. This is achieved by setting the weights of the recurrent connections from the output units to the state units to form an identity matrix and connecting each state unit with itself with a fixed weight (this weight affects the exponent of the averaging). The state vector thus constitutes the temporal context for the actions, and has two desirable properties: the state vectors at nearby points of time are similar, and state vectors corresponding to producing similar subsequences are similar as well. Because the memory in the Jordan network is based on an output-input feedback loop, this architecture is often encountered in the context of robot motor control and the modelling of proprioception. From the models of counting quoted in chapter 3, the model of Ahmad et al. (2002) used Jordan-like recurrence at multiple levels of the model architecture. The fact that the amount of

feedback in the Jordan network is dictated by the dimensionality of the output can however be viewed as a limitation. While the capacity of the Elman network can be easily adjusted by changing the number of units in the hidden layer, this is not the case with the Jordan network, for which adding or removing output units may have no reasonable interpretation (e.g. when outputs of the network correspond to the degrees of freedom of a robotic arm being controlled). Therefore, mixed architectures, which combine the concepts from both networks — recurrence in the hidden layer and output-input feedback — are also encountered (Stramandinoli, Marocco & Cangelosi, 2012).

Supervised training of simple recurrent neural networks can be achieved with a simple extension to the classical backpropagation algorithm (Rumelhart, Hinton & Williams, 1986). This technique, known as *backpropagation through time*, is based on the observation that, assuming that a finite period of time is considered, for every recurrent network a feed-forward network with identical behaviour can be constructed (Rumelhart et al., 1986, p. 354). In short, for every time step of the sequence being trained, a copy of every layer of the network is created, and the connections recurrent in the original network are turned into feed-forward connections that link consecutive copies of the network. The only necessary technical modification to the backpropagation algorithm is to ensure that all connection weights in the expanded network that correspond to the same connection in the original network have the same value. This constraint is easily met by updating the corresponding weights by the same value, which is based on the sum of the changes prescribed to the individual connections (Rumelhart et al., 1986, p. 355).

4.4.2 Continuous-time Neural Networks

Even in situations where the processing of temporal sequences is not involved, it is sometimes desirable to incorporate the temporal structure of the decision process in a neural network model. For this purpose, a class of neural network models known as the *firing rate models* (Shriki, Hansel & Sompolinsky, 2003; Yamashita

& Tani, 2008) can be employed. These models express the activity of neurons (or of ensembles of neurons) in terms of their mean firing rate, averaged over a short period of time (H. R. Wilson & Cowan, 1972; Hopfield, 1984).

Dynamics of a neural network in the firing rate model are described by a set of coupled differential equations, that specify how the activity of each unit changes over time. Assuming that the activity (firing rate) of the unit i is denoted as y_i , these equations have the form:

$$\frac{dy_i}{dt} = -\alpha y_i + \beta f_i \left(\sum_j w_{ij} y_j \right) + I_i(t) \quad (4.1)$$

where w_{ij} is the strength of the connection between units i and j , f_i is the activation function of the unit i , $I_i(t)$ is the external input to the unit i , and α and β are positive constants. The behaviour of a unit described by the equation 4.1 can be summarised as follows. The first term of the equation 4.1 is the *decay term* and ensures that, in the absence of external inputs (remaining part if the equation 4.1 equal to 0), the activity of the unit will tend to 0, independent of the initial state of that unit. If there is external input to the unit that has a constant value over time, the activity of the unit will approach asymptotically this value. Formally speaking, assuming that $I_i(t) = I_i$, it can be shown (Pineda, 1987), that the fixed points of the set of equations 4.1, denoted as \mathbf{y}^∞ , are given by the solution of the set of equations:

$$\alpha y_i^\infty = \beta f_i \left(\sum_j w_{ij} y_j^\infty \right) + I_i \quad (4.2)$$

If the neural network is feed-forward, the matrix of weights W is lower triangular, and therefore the solution to the set of equations 4.2 can be found using the substitution method. As Pineda (1987, p. 2229) points out, the propagation of activations in a discrete-time feed-forward artificial neural network is therefore equivalent to directly calculating \mathbf{y}^∞ for the situation when W is lower triangular. This has important implications in the context of training continuous-time neural networks: if the network is feed-forward, its training can be based on the standard backpropaga-

tion technique, using the values of the activations of the units after a stable state has been reached (see e.g. Verguts et al., 2005). Training continuous-time neural networks with recurrent connections is much more complicated and computationally expensive (cf. Yamashita & Tani, 2008). In the present work, only feed-forward networks in the firing rate model are considered.

Usually, the architecture of the network and the temporal structure of the inputs is complicated enough to make finding the solutions to the equations 4.1 analytically too tedious. A more practical way to determine the evolution of the activations of units y_i over time is therefore numerical integration. For that purpose, an appropriate numerical integration method of sufficient accuracy should be applied. Although studies exist which employed the very simple Euler method (Chen & Verguts, 2010, p. 238), because of its high inaccuracy and instability (Butcher, 2008), in the present study more sophisticated methods have been used.

One of the key benefits of modelling the temporal evolution of the network activity using the firing rate model is the possibility of simulating the chronometric data. This is commonly done by assuming that the response is given by the network when the activity of one of its output units reaches a predefined threshold value. The response time is therefore modelled ‘literally’ and expressed in integration steps. Examples of studies reviewed in chapter 3 that employed the firing rate model are Grossberg and Repin (2003), Verguts et al. (2005), Gevers, Verguts et al. (2006) and Chen and Verguts (2010).

4.4.3 Self-Organising Maps

A *Self-Organising Map* (often abbreviated as SOM, Kohonen, 1982, 1990) is a special type of an artificial neural network, that, in essence, transforms vectors in a high-dimensional input space to a low-dimensional output space. This transformation preserves, to a certain extent, the topological properties of the input space. In other words, vectors that are similar in the input space, are, in principle, transformed to vectors that are similar in the output space.

The properties of SOMs are inspired by results in neuroscience which suggest that various areas of the brains of higher animals employ representations that are topological in nature, what means that ‘a particular location of the neural response in the [cortical] map often directly corresponds to a specific modality and quality of sensory signal’ (Kohonen, 1990, p. 1465). A good example of such a ‘map’ are colour-selective cells found in the fourth visual areas (V4) of the monkey cortex (Zeki, 1980). SOMs provide a mathematical framework useful in the theoretical study of the principles that govern such brain representations.

A SOM consists of a number of units, each associated with a vector of weights \mathbf{w}_i , equal in the dimension to the number of dimensions of the SOM input space, N . Each unit is also assumed to have assigned fixed coordinates in the output space (usually low-dimensional, $M < N$), what defines the SOM *topology*. Upon the presentation of a vector $\mathbf{x} \in \mathbb{R}^N$ at the input to the network, the similarity between \mathbf{x} and the weight vectors of the map units \mathbf{w}_i is determined using a predefined metric (usually the Euclidean distance). The unit whose weight vector is the closest to \mathbf{x} is called the *best matching unit* (BMU). The coordinates of the BMU in the map output space can be taken directly as the output of the SOM, thus realising the aforementioned dimensionality reduction. Alternatively, a SOM may be used as a *sparse-coding* device, in which case the output from the network consists of the activations of all map units, calculated based on the distance between the weight vector of each unit and the input vector \mathbf{x} and an activation function of choice.

As evident from the above description, there is a great degree of flexibility with respect to the various parameters of a SOM. The most commonly encountered type of a SOM is simply a rectangular map, in which the units are located in the vertices of a square or triangular tiling of the underlying 2-dimensional Euclidean plane (see figure 5). In such a form, SOMs are particularly suitable for data visualisation. However, because of certain unwanted effects that appear in the very simple SOM topologies (e.g. the border effect, Li, Gasteiger & Zupan, 1993; Kohonen, 2001), more sophisticated solutions, like toroidal or spherical topologies, are sometimes also

applied (see for instance Nishio, Altaf-Ul-Amin, Kurokawa & Kanaya, 2006).

One of the attractive properties of SOMs is that the mapping from the input to the output space can be formed automatically, through unsupervised learning. The training algorithm involves repeated presentation of the training data vectors at the input to the network and gradual adjustment of the weight vectors associated with the map units. The crucial component of the weight update formula is the neighbourhood function Θ that specifies, for the given training vector \mathbf{x} , which units of the map are updated, and how strongly. Θ depends on the distance between the unit being adjusted and the BMU in the SOM output space, and changes throughout the training process. Initially, the neighbourhood encompasses most of the map units, but, as the training progresses, less and less of the units surrounding the BMU are affected. Assuming that $\text{BMU}(\mathbf{x})$ is the index of the BMU for the given input vector \mathbf{x} , the update equation can be written down as:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \Theta(i, \text{BMU}(\mathbf{x}), t) \cdot \alpha(t) \cdot (\mathbf{x} - \mathbf{w}_i(t)) \quad (4.3)$$

where t is the training iteration number and $\alpha(t)$ is the learning rate (which often decreases with t). Unfortunately, the proper values of the parameters of the training algorithm usually have to be found by trial and error. In particular, care has to be taken to ensure that a global order is established in the map in the early phase of training, and that it is not subsequently destroyed, e.g. through the use of a too aggressive learning rate or because of the neighbourhood shrinking too quickly (Kohonen, 1990, p. 1467).

The fact that the SOM training regime includes several compound parameters implies the necessity for having a means of assessment of the quality of the final result of the training. Multiple quantitative measures that attempt to express the goodness of a SOM exist (Kaski & Lagus, 1996). One of the most basic ones is the *average quantisation error*, which, for the given set of the data vectors X , is the mean distance between the data vectors and the weight vectors of their corresponding

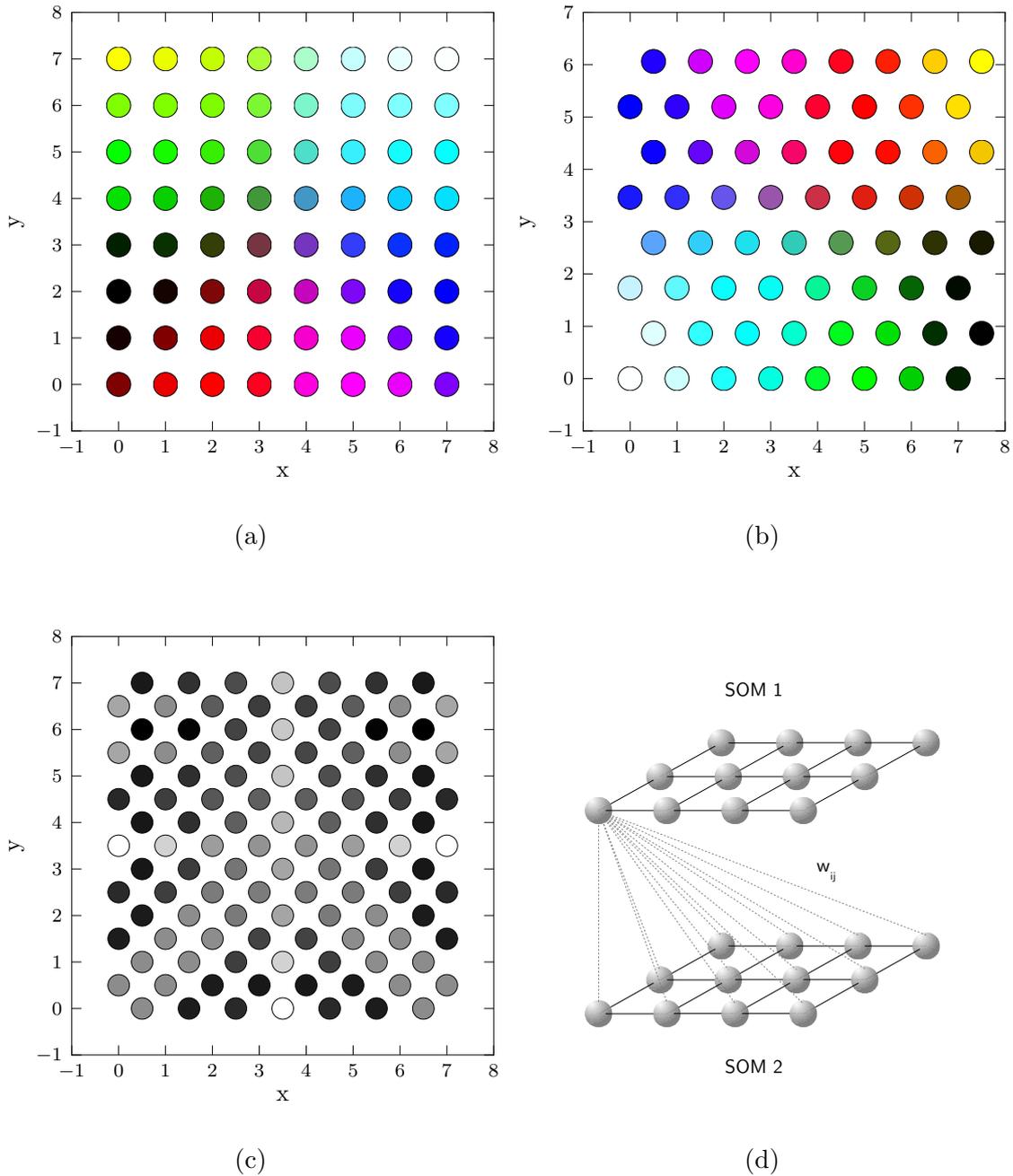


Figure 5: Example two-dimensional self-organising maps with rectangular 8×8 topology utilising square tiling (a) and triangular tiling (b). x and y are the coordinates of the SOM units in the SOM output space. The maps were trained to represent the three-dimensional additive red-green-blue colour space. Note that similar colours are mapped onto nearby areas of the maps. (c) shows the U-matrix of the SOM in figure (a). Note that the relatively brighter areas of the U-matrix correspond to the boundaries between the clusters of colours in the map. (d) shows a ‘stack’ of two 3×4 SOMs, in which all units of one map are connected with every unit of another map with Hebbian links. For clarity, only the set of links corresponding to one unit of the ‘top’ SOM is shown.

BMUS:

$$\text{AQE}(X) = \frac{1}{|X|} \sum_{\mathbf{x} \in X} \|\mathbf{x} - \mathbf{w}_{\text{BMU}(\mathbf{x})}\| \quad (4.4)$$

Average quantisation error is easy to compute and interpret, but it cannot be used as the sole measure of SOM quality, because very low values of the average quantisation error may be the result of over-fitting. A more sophisticated indicator of SOM goodness is the *topographic error* (Kiviluoto, 1996). It indicates the proportion of the data vectors from X for which the BMU and the unit second-best to the BMU are not adjacent in the SOM output topology:

$$\text{TE}(X) = \frac{1}{|X|} \sum_{\mathbf{x} \in X} u(\mathbf{x}) \quad (4.5)$$

where

$$u(\mathbf{x}) = \begin{cases} 1 & \text{if best- and second-best-matching units of } \mathbf{x} \text{ are non-adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Low values of the topographic error indicate good preservation of the topology of the input space. Note, however, that for this measure to be used, the definition of the SOM topology must allow for determining the adjacency of two units, what sometimes is not trivial. Another tool useful in the evaluation of a SOM is the *unified distance matrix* (U-matrix, Ultsch & Siemon, 1990). It is a method of SOM visualisation, in which the distances between the weight vectors (in the SOM input space) of SOM units adjacent in the output topology are a continuous variable displayed in the SOM output space (see figure 5c). U-matrices are particularly useful in identifying clusters in the input data. This is done either by visual inspection or using image analysis techniques. Similarly to the topographic error, U-matrices require the adjacency of the units to be well-defined in the SOM topology.

In cognitive modelling, SOMs find application most often as a sparse-coding mechanism, in which they encode a vector of variables using a population code. For example, the model of subitising proposed by Ahmad et al. (2002, pp. 179–180,

see chapter 3) uses a 1-dimensional SOM to represent the number magnitude, and a 2-dimensional one to represent number words as a map of verbal features.

SOMs can be integrated in a larger neural network model in various ways. It is common to use multiple maps that implement sparse coding of certain variables in a ‘stack’, connected ‘vertically’ with a set of modifiable connections (see Abidi & Ahmad, 1997; Ahmad et al., 2002; Morse, de Greeff, Belpeame & Cangelosi, 2010, and figure 5d). These connections implement a bi-directional translation from one population code to the other. Training of these connections is realised by an iterative presentation of corresponding input vectors to the SOMs and the adjustment of the connection weights using the Hebbian learning rule (Hebb, 1949).

It is possible to implement SOMs in the firing rate model (see Yamashita & Tani, 2008). In this case, the equations describing the activity of the SOM units have a similar form to the equation 4.1, but instead of (or in addition to) the weighted sum of the unit inputs, the argument of the activation function is the value of the similarity metric between the input vector and the unit’s weight vector. In order to minimise the computational cost of the simulation of such a network, if all possible values of the input vectors that will be presented to the SOM are known in advance, it is beneficial to pre-compute the SOM activation corresponding to every input vector and then pass it to the equation 4.1 as the external input $I_i(t)$.

4.5 Empirical Investigation of the Role of Gestures in Learning to Count in Children

In this section I review a behavioural study that focused on the investigation of the function of gesture in learning to count in children (Alibali & DiRusso, 1999). It is appropriate to delve into this particular research in more detail because the simulations of the first model presented in my thesis have been designed to be compatible with the experimental set-up used in this study.

4.5.1 Experiment Design

The particular focus of the experiments of Alibali and DiRusso (1999) was the distinction between two functions of the counting gestures: keeping track of the counted items and coordinating the recitation of the number words with the items being counted. In the study, the counting accuracy of 20 children, approximately five-year-old, was tested, in a variant of the HM task. The children counted sets of 7 to 17 chips arranged in a line with constant spaces between them. They were asked to count aloud; the instructions regarding gestures varied across seven experimental conditions:

- no instructions about touching or pointing;
- children instructed to point to (but not touch) the items being counted;
- children instructed to touch the items being counted;
- gesture prohibited — hands to rest clasped on the table;
- items pointed to by a puppet controlled by the experimenter;
- items touched by a puppet controlled by the experimenter;
- items pointed to by a puppet, with deliberate pointing errors;

The first condition allowed the researchers to investigate spontaneous counting gestures. The last condition was included in order to verify if in the passive gesture conditions the children were paying attention to the puppet's gestures. The dependent measure in the experiment was the children's counting accuracy, that is the number of trials in which they counted without any errors. The experimental design used by Alibali and DiRusso is illustrated in figure 6.

The principal scientific question of the study, that is whether the gestures are used only to keep track of counted items, was addressed by the above experimental design in the following way. First, based on earlier research, it was expected that the children would count more accurately when they were allowed to gesture than

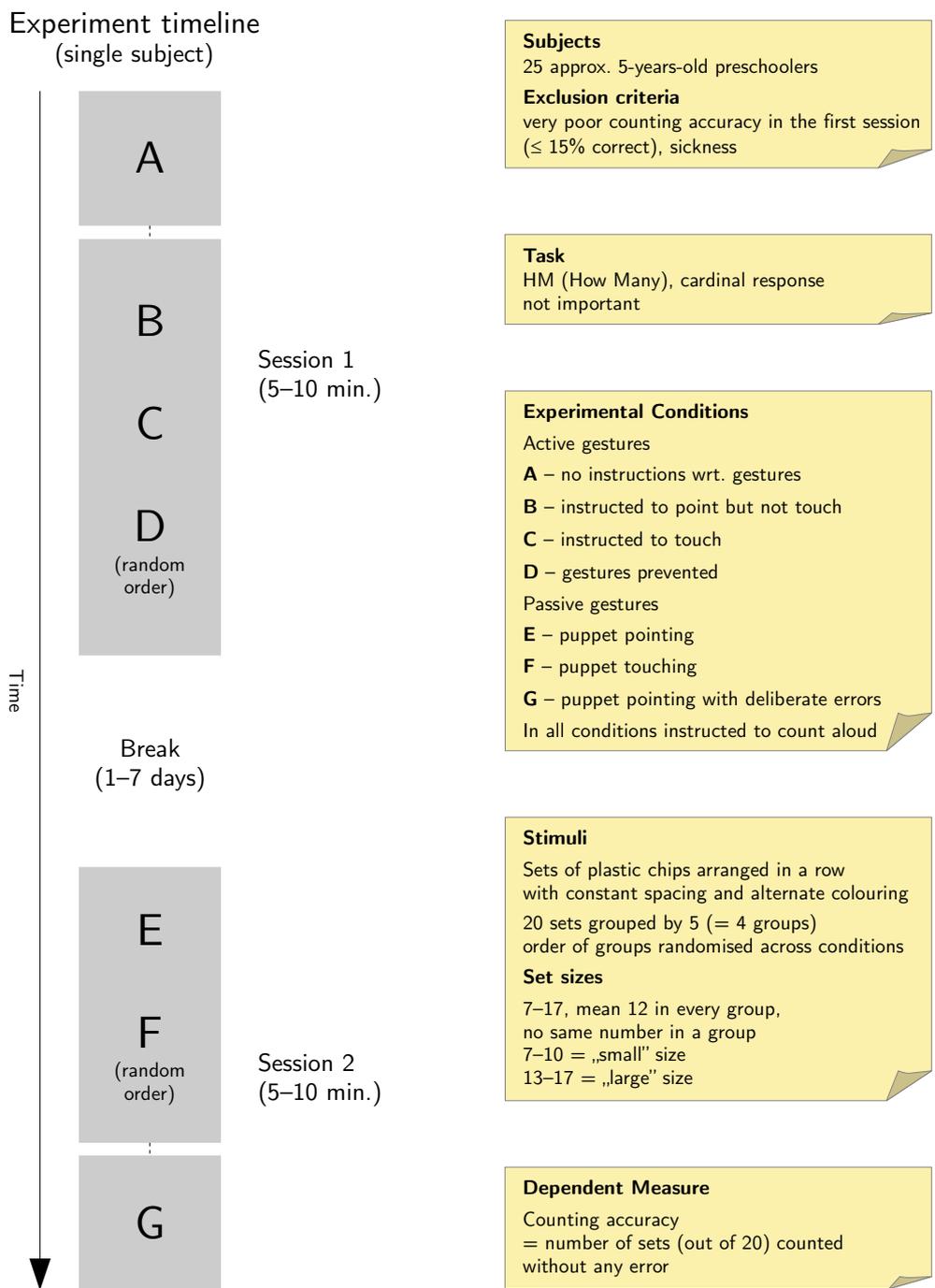


Figure 6: Experimental design employed by Alibali and DiRusso (1999).

when the gestures were prevented, supporting the ‘keeping track’ hypothesis (see e.g. Saxe & Kaplan, 1981). Second, should keeping track be the *only* function of the gestures, children would count in the puppet conditions at least as accurately as in the conditions where they gestured themselves, since the puppet’s pointing was always correct. The idea thus was to dissociate keeping track (which could be facilitated both by own and puppet’s gestures) from the coordination of recitation and tagging (which, the authors argued, could be facilitated solely by own gestures).

4.5.2 Counting Error Types

The evaluation of the children’s counting accuracy was based on the children’s gestures and on the correspondence between the gestures and the recited number words. Alibali and DiRusso (1999) distinguished six types of counting errors. Their error categorisation extended an earlier one, introduced by Gelman and Gallistel (1978). It is appropriate to repeat the definitions of the errors here (figure 7), since the same error classification is used in the present work.

It is important to note that the experimental condition affected the kinds of errors that could have occurred. For instance, in two of the puppet conditions it was not possible for children to double count or skip a chip, as the gesture made by the puppet was always correct and the experimenter took care to synchronise the puppet’s gesture with the child’s utterances (see Alibali & DiRusso, 1999, pp. 43).

4.5.3 Results

The results obtained by Alibali and DiRusso (1999) can be summarised as follows. First, a definite effect of the experimental condition was found. The children’s counting accuracy in the conditions with gestures was superior to the condition without them, as reported by earlier studies (e.g. Schaeffer et al., 1974; Saxe & Kaplan, 1981; Gelman & Meck, 1983). Moreover, the children counted more accurately in the conditions with touching than in the conditions with pointing (replicating Gel-

Code	Definition
Correct	The child assigns one number word to each chip, and uses the count words in the conventional order (or in an unconventional order that is used consistently across trials).
Partitioning errors	
Skip	The child does not assign a number word to a chip. ^a
Double count	The child assigns two or more number words to a particular chip. ^a
Coordination errors	
Continue	The child continues to say number words after the last chip has been indicated.
Stop short	The child does not assign a number word to the last chip (or last few chips) to be counted.
Other errors	
String error	The child uses the set of number words in an incorrect order that is used inconsistently across trials.
Distracted	The child is distracted from counting.

^a Skips or double counts of the *last* chip in an array were counted as coordination errors (Stop short and Continue).

Figure 7: Coding categories for children's counting performance. Reprinted from *Cognitive Development*, 14(1), Alibali, M. W. & DiRusso, A. A., *The function of gesture in learning to count: More than keeping track*, p. 44, Copyright 1999, with permission from Elsevier.

man & Meck, 1983), regardless of whether the gesture was performed by the child or the puppet. The data did not indicate however a statistically significant difference between the conditions with child's own gestures and those with puppet gestures, in terms of counting accuracy. These two groups of conditions differed however in terms of the frequency of the types of the counting errors that children made. The conditions with passive gestures were characterised by more frequent occurrence of coordination errors (continue and stop short) in comparison to the conditions with active gestures. Another effect found by Alibali and DiRusso was that of the set size. The children counted small sets (numbers of chips 7 to 10) more accurately than large sets (numbers of chips 13 to 17) across all experimental conditions. Finally, the study revealed the tendency of children to gesture spontaneously (pointing or touching, depending on the particular child) when no specific instructions have been given to them. However, in terms of counting accuracy, no statistical difference was found between the spontaneous and encouraged gestures. I will look at these results more closely in chapter 6, when comparing the behaviour of my model with the experimental data.

4.6 Plan of Work

The review of the ample experimental data available on mathematical cognition in chapter 2, and the related computational modelling efforts in chapter 3, followed by an overview of the relevant methodological aspects in chapter 4, set the scene for the presentation of the original research conducted in this thesis. The PhD research will investigate, with the aid of the tools provided by developmental cognitive robotics, two classes of embodied phenomena connected with learning to count.

The first class of the phenomena refers to a relatively short period of the development of human numerical knowledge, and focuses on the contribution of the counting gestures to learning to count. As reviewed in chapter 2, section 2.2.2, the effect of the counting gestures on counting accuracy in children is profound, and

various lines of experimental evidence suggest that such gestures are an important embodied cue, which aids, in some way, the acquisition of the conceptual understanding of counting. Inspired by this, the aim of the first modelling experiment, presented in chapters 5 and 6, will be to replicate this phenomenon in the iCub humanoid robot. I am going to propose an artificial neural network model that will make it possible for the robot to learn to count with and without gestures, and I will assess the obtained counting accuracy according to a methodology based on that used in the studies with children. By simulating learning to count in the robot I will try to answer the question if the proprioceptive information connected with the counting gestures represented in the form of values of joint angles that change over time can be helpful to an artificial learning agent (research question 2). Furthermore, inspired by some of the hypotheses about the mechanism of the contribution of the counting gestures put forward by the psychologists (cf. section 2.2.2) I will also conduct simulations aimed at investigating the significance of the simultaneous correspondence of the counting gestures to the number words being recited in the temporal domain, and to the items being enumerated in the spatial domain (research question 3). Although the model has been designed primarily with the above two research questions in mind, it turns out the approach adopted makes it possible to address also the research question 1.

The second group of experiments (chapters 7 and 8) assumes a broader perspective in terms of the period of the development that is being looked at. More specifically, using an analogous methodology of simulating embodied phenomena with the aid of an artificial body provided by the iCub humanoid robot platform, I will investigate if it is possible to replicate the acquisition of spatial-numerical associations in an artificial learning agent as the result of systematic spatial biases present in the development environment (research question 4). The evidence I quoted in chapter 2, section, 2.2.3, suggests that the degree to which humans internally link the representations of numbers and space is striking. Although quite a lot is known about spatial-numerical associations in general, the nature of the mech-

anisms behind their ontogeny is still an open research question (cf. section 2.3.3). In my simulation experiments I am going to extend the previous efforts in the computational modelling of the interactions between numbers and space (reviewed in section 3.4), by setting up a developmental process inspired by the progress of the acquisition of numerical knowledge by children. This process will involve the construction of spatial representations, as well as a simulation of the aforementioned spatial biases in learning to count, with the use of the iCub robot. In the subsequent behavioural simulations I will assess if the simulated developmental sequence leads to the manifestation of the association of numbers and space in the robot. This will be achieved with the classical tasks used in the same context for humans, such as timed number comparison, parity judgement and visual target detection (see section 2.2.3).

The motivations behind the focus on embodied numerical cognition are twofold. From the point of view of developmental robotics, the investigation of mathematical cognition is attractive because numerical knowledge in general, and the concept of number (the understanding of which is believed to be acquired in the course of learning to count) in particular, can be put forward as prime examples of abstract concepts. The question how abstract concepts can be represented and acquired by an artificial learning agent is still open, yet it is of fundamental importance for the artificial intelligence research (Barsalou, 1999). The replication of the acquisition of numerical knowledge in robots, in a way that is plausible from the point of view of cognitive science, can therefore be expected to provide useful pieces of information that will contribute toward solving this puzzle. In turn, from the point of view of cognitive science, numerical thinking is so widespread in our lives and so crucial for many of its aspects that hopefully it is not necessary to convince the reader about the importance of the investigation of the mechanisms which enable humans to first of all come up with the mathematical ideas, and then to reason, understand, and create with the use of them (Campbell, 2005). As discussed in chapter 3, computational modelling is a useful tool often employed in the research on cognition, and

section 4.2 of this chapter introduced cognitive developmental robotics as a methodology that elegantly supports computational modelling of the embodied phenomena. Considering the ample evidence for the embodied roots of mathematical thinking quoted in chapter 2, it is somewhat surprising that, to the best of my knowledge, at the time of writing no efforts to employ cognitive robotics as an aid in the study of mathematical cognition have been made. Hopefully, the efforts in this direction undertaken in the present thesis will pave the way for the future research, which will further our understanding of the nature of our mathematical thinking.

Part II

Neuro-Robotic Model of Learning to Count

Chapter 5

Model Overview

In this chapter I introduce a novel neuro-robotic model of learning to count, which purpose is the modelling of the contribution of the counting gestures to the acquisition of the counting skill, with a view to answering the first three of the research questions posed at the beginning of this thesis. The chapter starts with the discussion of the assumptions that have driven the design of the model. These assumptions result from the aims of the study and from the analysis of the current behavioural findings concerning the phenomena being modelled as well as of the previous attempts to model learning to count, reviewed earlier. Subsequently, the architecture of the model is described in detail, together with all its representational elements and appropriate training procedures. This includes an explanation of how the proprioceptive signal connected with the counting gestures is obtained using the iCub humanoid robot. The chapter concludes with a discussion of the proposed architecture in the context of the earlier models of counting.

5.1 Model Design Assumptions

Consistent with the aims of the conducted modelling experiment, the following assumptions have been made when designing the neuro-robotic model of learning to count with gestures, described in this chapter:

Design Assumption 1. Similarly to the other cognitive models formulated in

the context of mathematical cognition (cf. chapter 3), the model is implemented in the Parallel Distributed Processing (PDP, or artificial neural networks) framework (Rumelhart & McClelland, 1986). It is important to note however that the proposed neural network is not intended to be a neurophysiological model of any particular area of the brain at some specific stage of human development. Rather, the connectionist framework is used here as a tool for learning-based, sub-symbolic data analysis. The reasoning behind such an approach is that if a relatively simple learning agent, controlled by an artificial neural network, is able to exploit to its advantage the information carried by the gestures, it is highly likely that much more elaborate learning system, the human brain, is able to do the same. In addition, the relative simplicity of a PDP model makes it possible to investigate which properties of the gesture signal are important in the context of counting, by analysing how the model works.

Design Assumption 2. The model should be designed in a way that makes it possible to compare its counting behaviour with that of children. This implies that the model needs to be tested according to an experimental protocol that is as compatible as possible with that applied to children. In particular, the study by Alibali and DiRusso (1999), who investigated the function of gestures in learning to count, has been chosen to be the main source of the reference behavioural data (see chapter 4).

Design Assumption 3. At the centre of interest in the present work is the modelling of counting, which is an inherently sequential process. This is in contrast to subitising, which is an instantaneous apprehension of numerosity (cf. chapter 2). Because of the necessary simplifications that have to be made when modelling a real-world counting set-up, subitising may be a strong attractor in the model training dynamics. Care must be taken therefore to assure, through the inductive bias present in the design of the neural network, of the representational elements, and of the training regime, that the model employs a process of enumeration, rather than recognition, in order to determine the number of items in the counted set (Gordon

& Desjardins, 1995). Among others, this means that the proposed artificial neural network must be capable of processing temporal sequences. On the other hand, the amount of the inductive bias should be minimal, if the model is to provide useful insight into the investigated phenomenon. For example, the model should not be complicated beyond what is necessary, as this may result in it being ‘over-fit’ to the task. The above considerations express one of the major challenges in modelling counting.

The above design assumptions are reflected in the choices regarding the model architecture and the employed representations, which are discussed in detail below.

5.2 Model Architecture

The proposed cognitive model of learning to count is a recurrent artificial neural network, which combines the elements typical for the Elman and Jordan networks (see figure 8). The model is designed to accept visual stimuli and proprioceptive gesture information as inputs and to produce a sequence of number words and counting gestures as output. The response of the model is triggered by a separate, dedicated input. As indicated in figure 8, various components of the model are *optional*, that is they may be included in the model or not, making it possible to test it in various experimental conditions (e.g. counting with and without gestures), what will be the principal tool in addressing the posed research questions. In addition, this enables the process of the model development to be more in line with the experimental findings (this point is elaborated in section 5.3). The components of the model and employed representations are described below.

5.2.1 Vision

Visual input to the model is represented by a saliency map, which can be considered a simple model of a retina. Such a map consists of several units, each corresponding to

one spatial location in the robot’s visual frame of reference, which may be activated or not depending on the presence of an object at this particular spatial location. Such saliency maps are commonly employed to model visual stimuli (Itti & Koch, 2001), noteworthily also in the context of mathematical cognition (e.g. Dehaene & Changeux, 1993; Peterson & Simon, 2000; Ahmad et al., 2002; Verguts & Fias, 2004), and are hypothesised to exist in the brain (Koch & Ullman, 1985).

Consistent with the design assumption 3, the visual stimuli representation should be chosen in a way that discourages the neural network from inferring the cardinality of the set instantaneously, in one time step of the simulation, as this would correspond to subitising rather than counting. There are at least two ways in which the exact number of items in a saliency map could be deduced by a simple feed-forward neural network. The first way is by calculating the sum of activations of all units in the visual input. In order to eliminate this possibility, in the employed implementation of the saliency map the activations of its units are normalised so that their sum is always equal to 1, regardless of the number of items in the represented set. For example, a collection of 5 items activates 5 units in the map, each with the activation value equal to $1/5 = 0.2$ (remaining units are inactive, i.e. their activation value is equal to 0). This is illustrated in figure 9.

The second way a neural network could easily infer the cardinality of a set is by exploiting a correlation between the number of items and their spatial locations in the visual input, should such correlation exist. This would be the case, for example, if the spatial arrangement of a collection was always the same for a particular numerosity. Consequently, the spatial locations of the items have to be randomised between trials, and the number of units in the visual input layer has to be large enough to allow such randomisation to a sufficient degree, even for the largest considered numbers. In order to achieve this, in the simulations reported in this thesis, the number of units in the visual input is always set to be twice the size of the biggest considered collection ($N_{VI} = 2N_{VO}$). Such an assumption also ensures that a more numerous set can be arranged in a bigger number of configurations than a

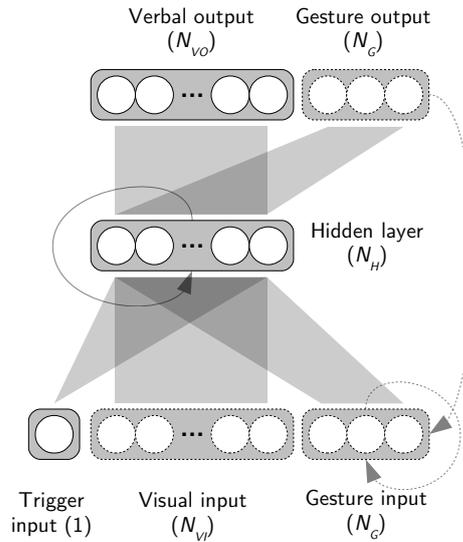


Figure 8: Architecture of the artificial neural network model of the contribution of gestures to learning to count. Neural activity is propagated from the bottom to the top. Grey areas represent all-to-all connections, and arrows represent recurrent connections. Number of units in each layer is specified in parentheses next to the layer name. The layers outlined with dotted lines and their corresponding connections are optional, i.e. can be present or not depending on the desired model configuration.



Figure 9: Visual input representation example. Locations of the counted items which are assumed to be arranged in a row (top) determine which units of the visual input layer (consisting here of $N_{VI} = 10$ units) are activated (middle). The activation values of the visual input units are normalised so that the total activation over the layer equals 1 (bottom).

less numerous set over the entire considered range of set sizes (this would not be the case if numerosities greater than $\frac{N_{VL}}{2}$ were considered).

An important parameter of the visual saliency map is the number of its dimensions. A three-dimensional map, in which each unit encodes a physical location in the peripheral space of the robot, would probably correspond most closely to how humans perceive space around them. Considering however that during learning to count children most often are exposed to sets of items arranged in two dimensions (e.g. pictures on a book page, toys lying down on the floor), a two-dimensional map should model such situation sufficiently well. If one assumes that only items arranged in a row are considered, a one-dimensional map is also a viable option. Decision regarding the dimensionality of the saliency map is important, because it is reasonable to expect that it will affect the difficulty of the training of the neural network. As discussed in chapter 2, counting involves navigation through the set of items along a path that ensures a correct result of the enumeration. The larger the dimensionality of the saliency map, the more difficult it is to construct such a path, because the number of possible arrangements of items increases substantially with the number of units in the map. It is an established fact that the arrangement of the items affects how difficult counting is for children (Beckwith & Restle, 1966), therefore this parameter is controlled for in the experimental studies by consistently using a simple arrangement, often a row (e.g. Alibali & DiRusso, 1999; Graham, 1999; Le Corre et al., 2006). An analogous approach is adopted in the simulation experiments conducted in this thesis by employing a one-dimensional saliency map.

5.2.2 Gestures

The representation of the counting gestures is the crucial element of the proposed model of learning to count. There are at least two approaches to represent the proprioceptive information connected with the gestures that can be easily incorporated in the neural network architecture shown in figure 8. The first one is based on the encoding of the location being pointed to, expressed for instance in the same

frame of reference as the visual input. Such a representation would consist of a cluster of units of the same shape as the visual representation layer, and an activated unit in this cluster would indicate the spatial location that is currently pointed at. Executing the pointing gesture would correspond to shifting the activation from one unit to another in this layer. Such a rather abstract representation of gestures was used in the model of counting of Ahmad et al. (2002). Another approach, which would not be possible without adopting the cognitive robotics methodology, is to provide the neural network with the values of the angles of the joints of the robot arm, with which the pointing is performed. Here, the appropriate changes in the angles of the arm joints are fed to the neural network as the robot performs the counting gestures.

The proposed neural network architecture can be used with both approaches described above. However, as the way the research question 2 has been formulated reveals, in the present work the focus will be put on the latter representation. This choice is motivated by the fact that, arguably, this resembles more closely the proprioceptive information available to a child while it performs the counting gestures. In mammals, the angles of the joints of the limbs are (together with velocity, acceleration, and jerk) among the primitive information collected by the estuproprrioceptive sensor neurons located directly in the muscles. This information is then passed on to the spinal cord and, further, the brain (Arbib, 2002a). Note that since velocity and acceleration are simply derivatives of the joint angle values, they do not carry much additional data from the point of view of an artificial neural network. Providing the model of counting directly with the angles of the joints of a pointing robot arm leaves the model the freedom to find a suitable way of making use of this information, releasing the researcher from making arbitrary assumptions about how the raw signal from the muscles is actually processed. Introducing such assumptions could bias the way in which the proprioceptive signal is exploited by the neural network. Therefore, by using the joint angles, it is possible to investigate if the gesture information is useful to facilitate learning to count even without ex-

plicit pre-processing. Consequently, in the experiments presented in this thesis, the representation of gestures based on the joint angle values has been used.

Independently of the choice of the method of representation, there are two further issues about the gestures that need to be addressed. The first one is connected with the relation between the proprioceptive signal and the artificial neural network. As indicated in figure 8, in the proposed model the output-input feedback connected with the gesture layer, typical for the Jordan network architecture, is optional. In other words, the gesture may either be *produced* by the neural network, or it can be available to it as an external *input signal*. This contributes to the flexibility of the model in the sense of the number of different scenarios in which it can be tested. Although the ‘gesture production’ set-up may at first seem to be the more natural option, various experimental findings justify the other approach as well. First, children’s counting accuracy improves even when they do not point to the counted items themselves, but only observe somebody else’s counting gestures (passive gesture conditions in Alibali & DiRusso, 1999). In this case the gestures are for the children, in a sense, an ‘external input signal’. Second, even in children’s own counting, the one-one correspondence principle appears initially in the gestures and only subsequently it is transferred to speech (Graham, 1999). Finally, from a more general point of view, providing the gesture signal as an input to the neural network makes it possible to separate the effects of learning the correct gestural activity from the influence of the latter on learning to count.

The second issue connected with the representation of the gestures is how to encode the performing of no gesture (e.g. ‘not pointing’). This is necessary for instance to appropriately model the termination of the counting procedure. In the case of both representations of gestures discussed earlier, several choices are possible. For instance, Ahmad et al. (2002) included in their gesture encoding layer an additional unit, which, when activated, represented the absence of a pointing action. An alternative, which would not require artificially extending the network, would be to treat the state in which all units in the gesture layer are deactivated

as representing ‘not pointing’. For the joint angle representation, a special ‘rest’ posture of the robot arm can be introduced, to which the arm should go once the counting is completed. Another option is to assume that at the end of counting the robot should leave the arm pointing to the last indicated item. The decision regarding the representation of ‘not pointing’ is not without importance. This should be evident when the configuration of the model with the gesture acting as an input to the network is considered. In this case, should the representation of ‘not pointing’ be distinctly different from the representations of the counting gestures (e.g. a dedicated neural unit), the contribution of the gestures to counting would most likely be reduced to indicating the end of the procedure. The model could disregard the actual counting gesture, and simply learn to recognise the final posture as a trigger for suppressing further verbal output. While in principle this could allow the model to achieve high counting accuracy, such a behaviour would not be in agreement with the experimental data which suggest the importance of the entire counting gesture and not just of its termination (cf. chapter 2). As the result of these considerations, in the proposed model of counting, after all items to be counted have been indicated, the robot arm is assumed to remain pointing to the last item. Since under this assumption there is no unique representation of ‘not pointing’ (the final arm posture is different for different arrangements of the items), the model is more likely to consider the entire gesture trajectory.

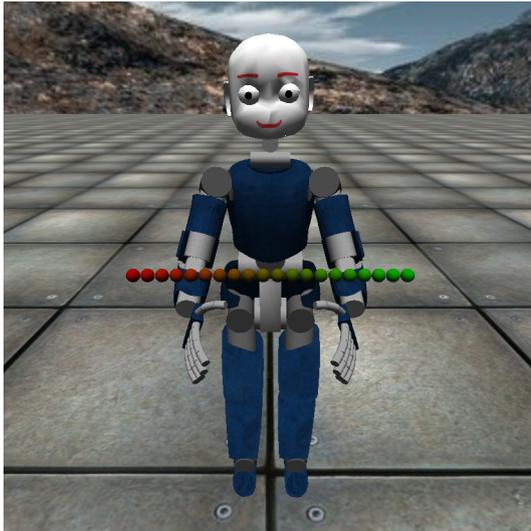
The representation of the counting gestures based on the joint angles, which served as the source of the target data for the proposed neural network, was constructed using the iCub humanoid robot (see chapter 4). Pointing gestures, intended to imitate those made by children, were performed using the robot’s right arm kinematic chain. Six degrees of freedom of this kinematic chain were used, namely torso yaw and pitch, and the first four joints of the robot arm, that is shoulder and elbow. The torso roll angle has been locked in order to eliminate unnaturally-looking body postures. Since in all simulations presented in this thesis it has been assumed that the visual input representation consists of 20 spatial locations arranged in a

horizontal row (the choice of $N_{VI} = 20$ is justified in section 5.2.3 below), the robot was commanded, using the Cartesian interface, to ‘point’ to 20 locations in front of it, which were assumed to correspond to the 20 spatial locations represented by the units of the visual input layer. These target locations of the robot arm’s end effector (the palm), intended to resemble the pointing gesture, were uniformly distributed on a horizontal line, placed 30cm in front of the robot, 10cm above its hip and spanned from 20cm left to 20cm right (see figure 10a). After the robot moved the arm to each of these locations (figure 10b), the joint angles of the arm were recorded and normalised according to the limits of the arm joints, yielding an initial proprioceptive representation of the counting gestures. The collected trajectories of the joint angles are shown in figure 11a.

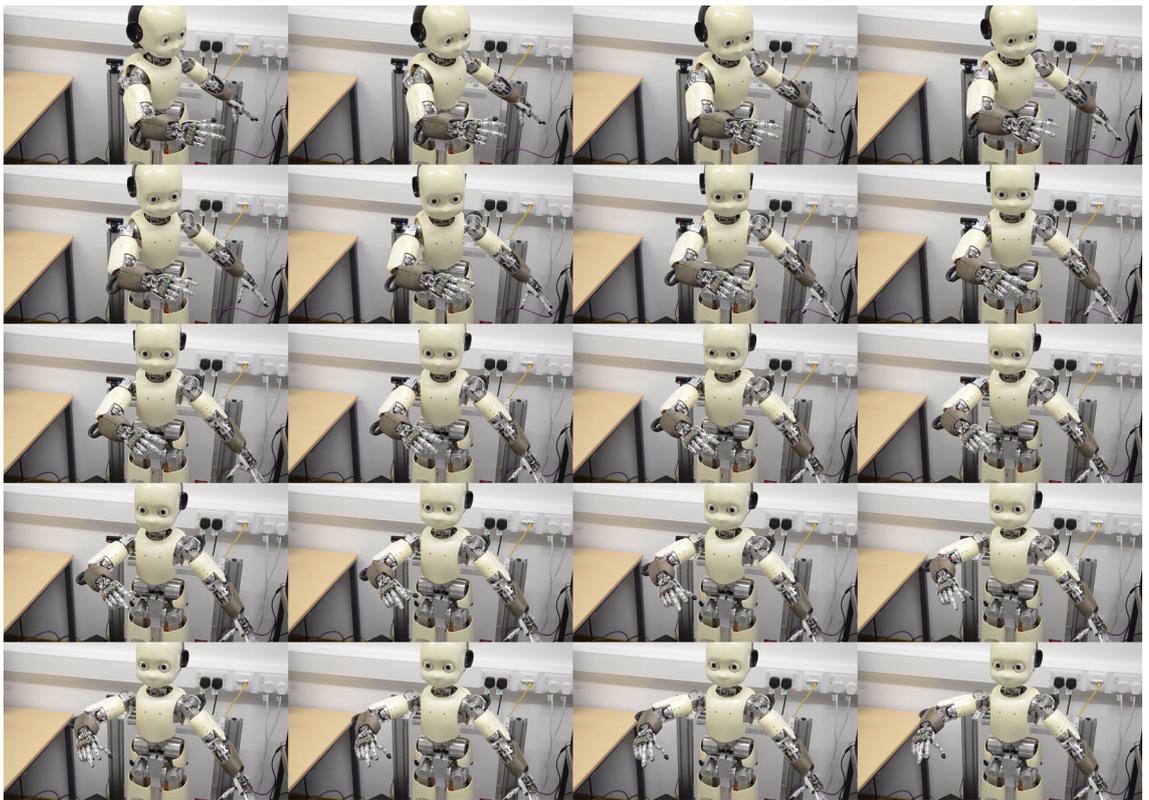
The values of the joint angles collected for the 20 considered locations were subsequently examined using the Principal Component Analysis, which revealed that the three strongest principal components carry more than 90% of the total ‘statistical energy’ of the original data. Moreover, a regression model based on these three principal components explained more than 99% of the variance of the joint angle values. Therefore, the dimensionality of the proprioceptive signal was reduced by assuming the three strongest principal components to act as the final representation of the counting gestures (yielding $N_G = 3$). Figure 11b shows the resulting trajectories of the three proprioceptive units for the 20 considered spatial locations. The applied procedure is a standard technique of pre-processing information before it is presented to an artificial neural network for training (LeCun, Bottou, Orr & Müller, 1998). It does not affect the ability of the network to learn from the data, but speeds this process up.

5.2.3 Speech and Number Words

As with the gesture representation, several coding schemes for the number words production are supported by the proposed model architecture. In previous works, random, uncorrelated states (Amit, 1988), binary representations intended to reflect

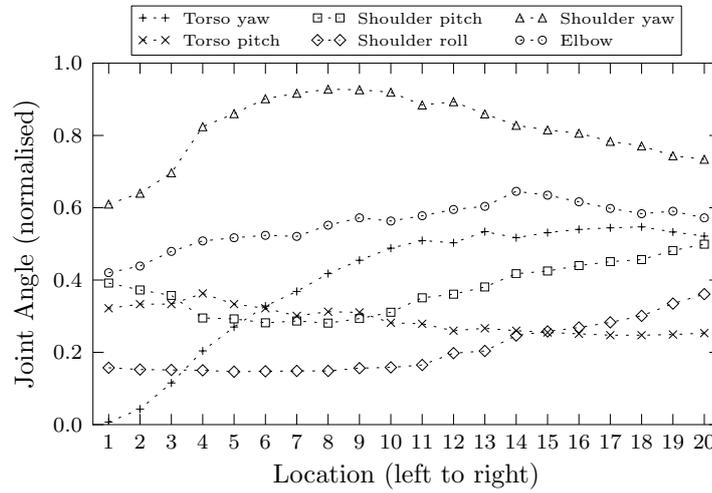


(a)

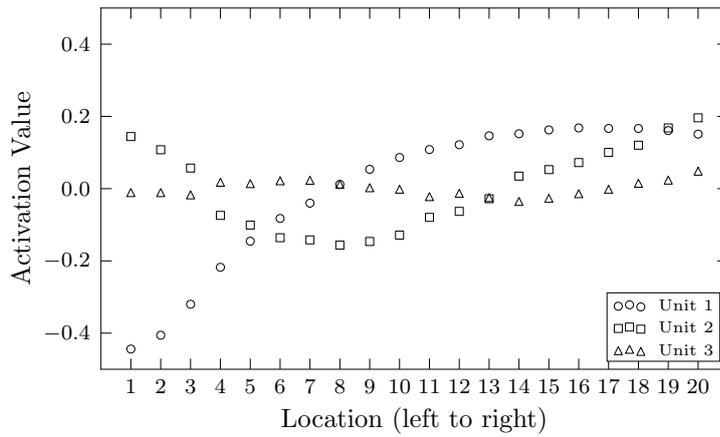


(b)

Figure 10: Counting gesture production using the iCub robot. (a) 20 target locations of the robot arm end effector visualised in the iCub simulation software. (b) iCub robot performing the ‘pointing’ to each of the 20 locations.



(a)



(b)

Figure 11: Counting gesture joint angles. (a) Recorded angles of the joints of the robot, normalised with respect to the joints limits. (b) The final representation of the proprioceptive gesture information obtained from the three strongest principal components of the data shown in (a).

similarity between the number words (Ma & Hirai, 1989) and self-organising maps based on the phonetic features of the number words (Ahmad et al., 2002) have been employed for this purpose. The most general method of representing number words in an artificial neural network is perhaps using one-hot encoding (used for instance by Verguts & Fias, 2004). In this coding scheme, every unit of a neural network layer represents one number word, and at most one unit in the layer is permitted to be active at any time. The number of the units in the layer (N_{VO}) must therefore be equal to the length of the count list the model is assumed to produce. A characteristic feature of this encoding is that the vectors representing the number words are orthogonal and therefore it is not relevant which particular unit in the neural network layer represents which number word. For convenience, it may thus be assumed that the adjacent units represent consecutive number words, without loss of generality.

Similarly to the gestures, the possibility of ‘not producing any words’ has to be considered also, in order to account for the end of counting in an elegant way. The same options as for the abstract gesture representation are available here, i.e. adding a special ‘silence’ output unit or deactivating all units in the layer. In the present work, the latter approach is used.

Once an output vector representing a number word is produced by the artificial neural network, a method of determining which word has actually been uttered is necessary, because of the continuous nature of the output of the model neurons. In the simulations reported in this thesis, when the model of counting is tested, the number word produced at each time step of the simulation is determined using nearest-neighbour classification (Cover & Hart, 1967). That is, the euclidean distance between the actual output of the artificial neural network and all prototype vectors representing the considered number words (and silence) are calculated, and the vector closest to the actual network output is assumed to be the produced utterance.

As indicated in sections 5.2.1 and 5.2.2, the choice of the length of the count list

the model should produce (which in case of one-hot coding is equivalent to N_{VO}) is linked with the visual and proprioceptive representations. In order to preserve the desired combinatorial properties of the realisable arrangements of items for the considered set sizes, N_{VI} should be equal to $2 \cdot N_{VO}$. N_{VI} in turn, determines the amount of data that have to be collected using the robot in order to construct the proprioceptive signal (and, in practice, it is limited by the precision of the robot). Keeping in mind the performance considerations (as the number of possible arrangements of objects grows exponentially with N_{VI} , excessive N_{VI} could imply impractically large amount of training necessary to achieve satisfactory performance of the model), as well as the fact that, in most languages, for numbers above 10 a regular structure of the number words appears, which may have an impact on learning to count (what is not being modelled at present), the simulations conducted within the scope of this thesis assume $N_{VO} = 10$, and therefore $N_{VI} = 20$. Simulation of learning a longer count list is of course a possibility in the proposed network architecture. However, it is likely that in such case the phonetic structure of the number words would have to be taken into account when choosing the verbal output representation.

In addition to the verbal output, the proposed model of learning to count has also a special one-unit input called the trigger input (see figure 8). It has been included for practical reasons, and its role is to indicate to the robot when the counting process should start. Activating this input represents asking a child a question (e.g. ‘How many?’) which according to the experimental protocol should encourage counting.

5.3 Training of the Model

Consistent with the behavioural data, according to which children are able to recite the number words quite well before they start learning to count items (see e.g. Gelman, 1980; Le Corre et al., 2006), the development of the proposed model proceeds in two stages. The first stage is intended to equip the neural network with an ability

to recite correctly a sequence of number words. Subsequently, in the second stage, the model is trained to count the items presented to its visual input. Introduction of such a distinction will make it possible to look for the answer to the research question 1 in one of the simulations described in chapter 6.

5.3.1 Preliminary Skill — Learning the Count List

The preliminary skill to be acquired by the model in the first stage of the training is defined as being able to output a pre-defined sequence of number words and then to remain silent for a period of time, in response to the trigger input. At this stage of the training none of the optional components of the model are present, i.e. the only input available to the model is the trigger input, and only the verbal output is present (see figure 12). Assuming the model is trained to recite N_{VO} number words, the training data set consists of two sequences, each $2N_{VO}$ time steps long:

- the first sequence, with trigger input equal to 0 and all target outputs in the ‘silence’ configuration at every time step, trains the model to remain silent when the trigger input is deactivated;
- the second sequence, with trigger input equal to 1, trains the model to recite the N_{VO} number words throughout the first N_{VO} time steps of the sequence, and to remain silent during the remaining N_{VO} time steps;

The prolonged period of silence following the recitation of the count list has been introduced in order to prevent any potential unwanted effects that may appear for the largest numbers when training the network to count in the second stage. Should the simulation sequence be equal in length to the biggest considered number, the latter would be subject to a different treatment than the lower numbers, in the sense that after counting up to the largest number the neural network would not have to enter a stable ‘silence’ state. The length of the period of silence assumed to follow counting (equal to N_{VO}) was chosen arbitrarily.

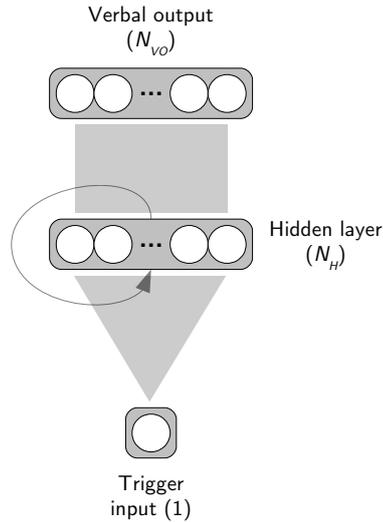


Figure 12: Architecture of the neural network in the first stage of the training. Note that this represents a subset of the neural network shown in figure 8, with no optional components.

Before the training of the model commences, the adjustable weights of the neural network are initialised randomly and drawn from a uniform distribution with mean 0 and standard deviation equal to the reciprocal of the square root of the fan-in of the node at the receiving side of the connection (LeCun et al., 1998). The first stage of the training is performed using the RProp– algorithm (Igel & Hüsken, 2003). In the simulations reported in chapter 6 the training typically lasts for 700 epochs with the initial learning rate of 0.1 (the learning rate is allowed to vary between 10^{-5} and 0.1). The number of the units in the hidden layer N_H is usually a parameter of the model.

The success of the first stage of the training is determined based on the correctness of the output produced by the model. If the nearest-neighbour classification of the network output is the same as that of the target output at every time step for both sequences in the test data set (which in this stage of the training is identical with the training data set), the preliminary stage can be deemed successful, and the training can proceed to the second stage. Should the preliminary stage not be successful, it would not make sense for the training to proceed.

5.3.2 Counting

After the first stage of the training is completed, the model obtained in the preliminary stage is extended by adding the desired optional components (cf. figure 8). The weights of the connections that need to be added (e.g. from the visual input layer to the hidden layer) are initialised using the same method as in the preliminary stage. Subsequently, the extended neural network is trained to produce a sequence of number words (and, optionally, gestures), the length of which is equal to the number of objects present in the visual input, just as a child would when counting the same set. As explained in section 5.2.1, in the training data set the spatial positions of the counted items must be randomised. The number of possible spatial arrangements of the objects grows exponentially with the number of units in the visual input. If one considers numbers up to 10, what in the model corresponds to the visual input with $N_{VI} = 20$ units, there are more than 600,000 possible spatial arrangements of the items. Already in this case it is impractical to create a training data set that would contain all possible combinations of locations of objects for all considered numbers. In order to alleviate this problem, the model is trained in an ‘on-line’ fashion, using small data sets that change after every training epoch. This makes it possible to use a different spatial arrangement of objects for a particular number in every epoch.

The training data sets for the second stage of the training are constructed as follows. For every number from the considered number range (up to N_{VO}), the representation of the visual input is composed by randomly choosing the locations of the items within the saliency map. The visual input fed to the model for a particular number remains unchanged throughout a simulation sequence. For every number, two sequences are included in the training data set. In the first sequence, the trigger input is deactivated and the target outputs are set to ‘silence’ (and, optionally, no gesture) throughout the whole duration of the simulation sequence. In the second sequence, the trigger input is activated and the target outputs contain the correct count list (and, optionally, gestures) that correspond to the current visual

input, followed by ‘silence’ (no gesture) until the end of the simulation sequence. If numbers from 0 to $N_{VO} = 10$ are considered, this results in 22 sequences per data set.

If the model configuration includes the optional elements connected with the gestures, the appropriate proprioceptive signal needs to be constructed, that corresponds to the given spatial arrangement of the objects being counted. In the simulations described in chapter 6, two types of gestures are considered: the ‘standard’ counting gestures, that have spatio-temporal character, and ‘rhythmic’ gestures, that have only the temporal aspect. How these two types of gestures are constructed based on the given arrangement of the items in the visual input is explained below.

5.3.2.1 Spatio-temporal Gestures

When constructing the spatio-temporal counting gestures, the locations of the objects in the visual input are considered in a particular order, e.g. from left to right. As described in section 5.2.2, every spatial location in the visual input layer has a corresponding vector of the activation values of the gesture representation units derived from the values of the joint angles of the robot arm pointing to this position. Assuming there are $K \leq N_{VO}$ objects in the counted set, for the time steps $0 \leq t < K$ of the simulation sequence, the target activation values of the gesture representation units are assumed to be equal to those representing a posture in which the robot points to the spatial location of the $(t + 1)$ -th object, in the considered order. For the remaining time steps $K \leq t < 2N_{VO}$, the activations of the proprioceptive units remain the same as they were for the last object (as discussed before, alternative ways of representing ‘no gesture’ are also possible). This process is illustrated in figure 13. Note that since the arm configuration is different for every location being pointed to, different spatial arrangements of items will yield different gesture signal, even if the collections are of the same size. A spatial correspondence exists therefore between the counted items and the gesture performed.

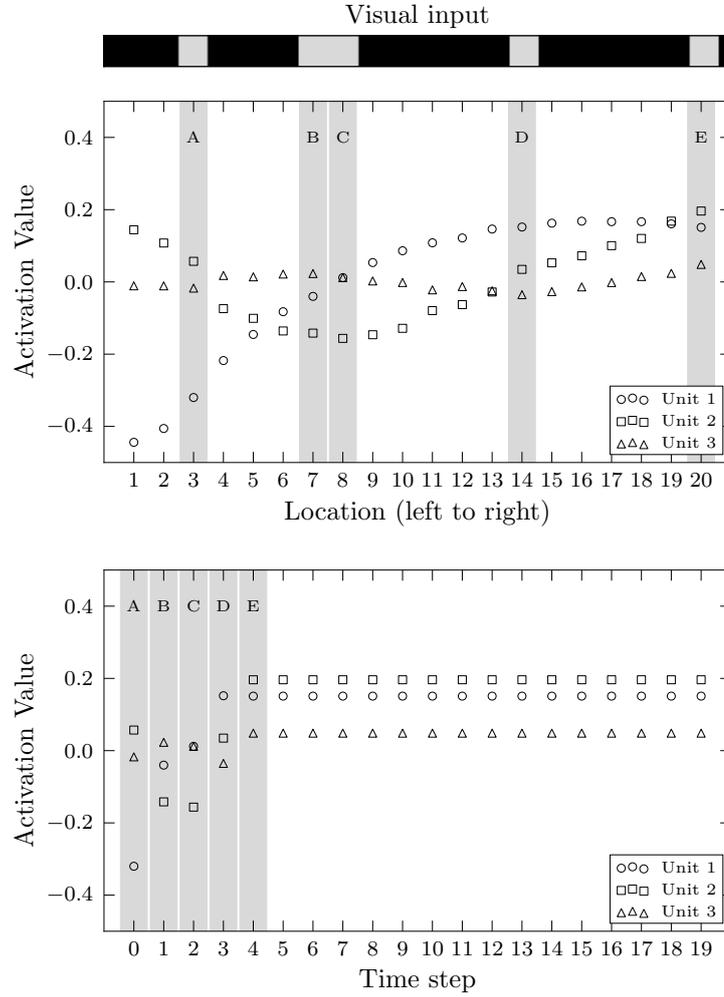


Figure 13: Spatio-temporal counting gesture construction example. It is assumed that $N_{VI} = 20$ and $N_{VO} = 10$. There are five items in the robot's visual input (top). The activated units in the visual input layer determine which arm postures are used to construct the gesture signal. The vectors corresponding to the occupied spatial locations are designated as A , B , C , D and E (centre). The counting gesture unfolds through the first five time steps of the simulation sequence, representing pointing to the items in the left-to-right order. After the gesture is completed (at time step 5) the robot arm remains in the posture corresponding to the last counted item until the end of the simulation sequence (bottom).

5.3.2.2 Rhythmic Gestures

Rhythmic gestures, employed in the simulation described in section 6.4, are constructed in the following way. Let l and r be two distinct spatial locations chosen from the $N_{VI} = 20$ locations represented by the units of the visual input layer. The choice of l and r affects the amplitude of the resulting rhythmic movement. For example, assuming $l = 1$ and $r = 20$, that correspond to the leftmost and rightmost spatial position respectively, yields the highest achievable movement amplitude. In turn, taking $l = 10$ and $r = 11$ is a way to obtain a movement with the smallest possible amplitude. If there are $K \leq N_{VO}$ objects in the set to be counted, the rhythmic gesture signal is constructed by taking, for the time steps $0 \leq t < K$ of the simulation sequence, the activation values of the proprioceptive units corresponding to the spatial positions l and r , interchangeably. For the remaining time steps of the simulation sequence $K \leq t < 2N_{VO}$, the target activations of the proprioceptive units remain unchanged with respect to those present for $t = K - 1$. This is illustrated in figure 14. Note, that any arrangement of K items in the visual input results in exactly the same gesture signal. As a consequence, in case of rhythmic gestures there is no spatial correspondence between the gestures and the counted items. The rhythmic gestures will be contrasted with the ‘normal’ counting gestures described above, in an attempt to answer the research question 3.

For the sequences in the training dataset where the trigger input is deactivated (for which counting does not occur), the activation values of the gesture inputs are set to 0, what corresponds to keeping the arm in a ‘neutral’ position. Since the training data set changes in every epoch, in the second stage the neural network is trained using the backpropagation through time algorithm with the network weights updated in an on-line fashion (LeCun et al., 1998). In simulations described in chapter 6, the training usually lasts for 4000 epochs and a constant learning rate of 0.005 is used.

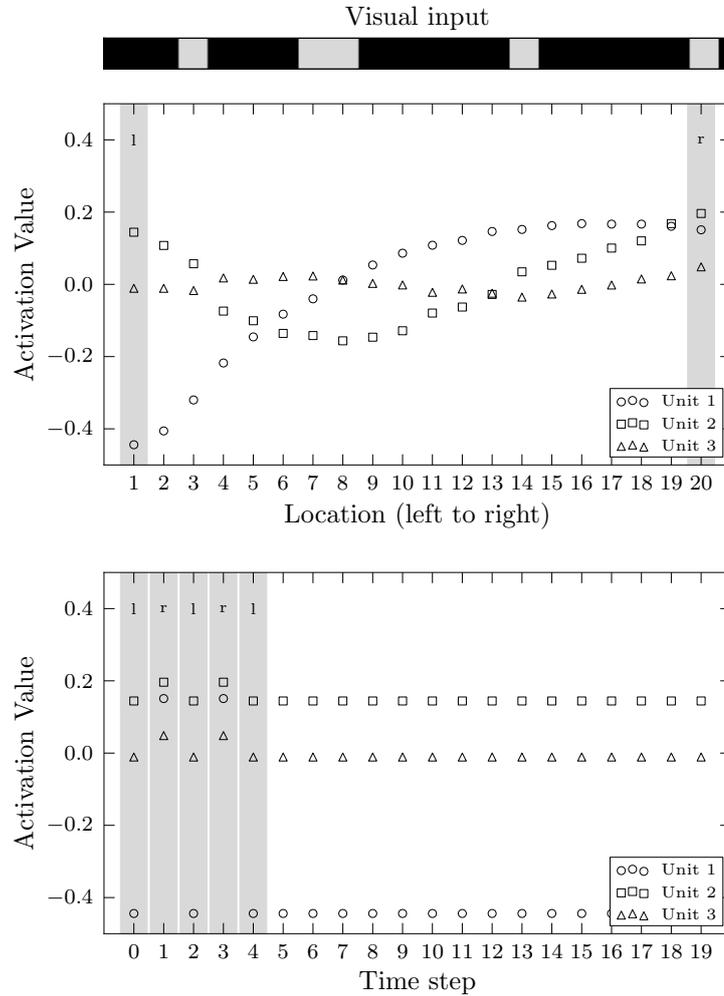


Figure 14: Rhythmic counting gesture construction example. It is assumed that $N_{VI} = 20$ and $N_{VO} = 10$. The same arrangement of five items in the visual input is used here as in figure 13 (top). $l = 1$ and $r = 20$ are assumed, therefore a gesture with maximum amplitude will be obtained (centre). The gesture is constructed by interchanging the vectors l and r five times (since there are five items). After the gesture is completed (at time step 5) the robot arm remains in the posture corresponding to the last beat (bottom).

5.4 Model Performance Evaluation

In line with the design assumption 2, the evaluation protocol in the conducted robotic simulations of counting aims to follow as closely as practically possible the one applied in the behavioural study by Alibali and DiRusso (1999), so that the results of both can be meaningfully compared. The ‘subjects’ of the simulated experiment are the instantiations of the neural network model architecture described earlier in this chapter. Inter-subject variability results from the random initialisation of the weights of the connections within the network, as well as from the stochasticity of the applied training algorithms. Typically³, the subjects are first trained to recite the sequence of number words as described in section 5.3.1, and the successful acquisition of this skill is required for a subject to be included in the remaining part of the experiment. Subsequently, several experimental conditions are simulated by extending the network with the desired optional components, training it appropriately to the experimental condition, and then evaluating it on a test dataset. Note that for a single subject the starting point of every experimental condition is a copy of the same neural network obtained in the preliminary training stage. This makes it possible to apply the repeated measures design in the statistical analysis of the results. The experimental set-up described above is illustrated in figure 15.

During the evaluation, the model is presented with the test stimuli, which consist exclusively of the arrangements of items that have never been shown to the model during training. Because of the small number of the possible spatial configurations of the objects for the smallest numbers of items, the test data set is determined prior to the commencement of the training, and the arrangements used in the test data set are prevented from being used throughout the training. The test data sets are constructed in the same way as described for the training data sets in section 5.3.2.

³Since the four experiments described in chapter 6 have different aims, there are minor differences in the details of the experimental set-ups between these simulations. The description herein refers specifically to the simulation experiment 3 presented in section 6.3, which is aimed directly at reproducing the results of Alibali and DiRusso (1999).

Experiment timeline (single subject)

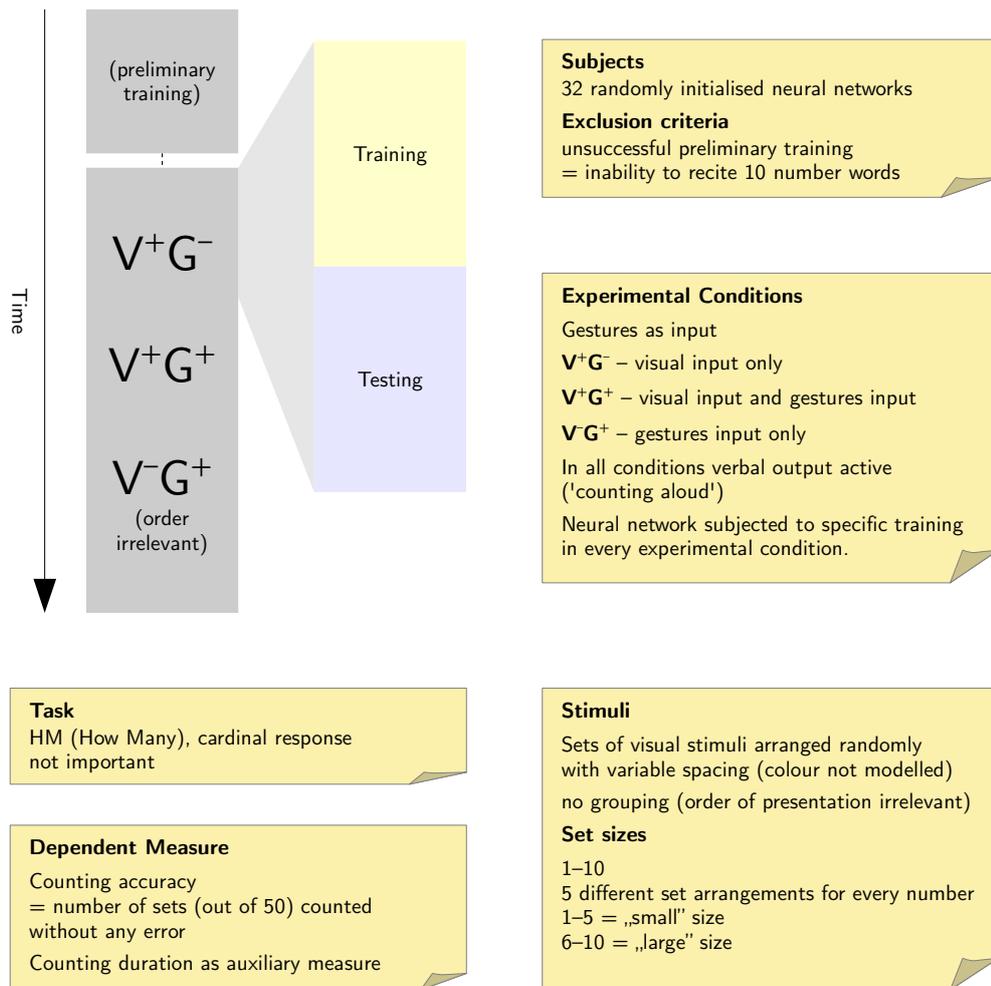
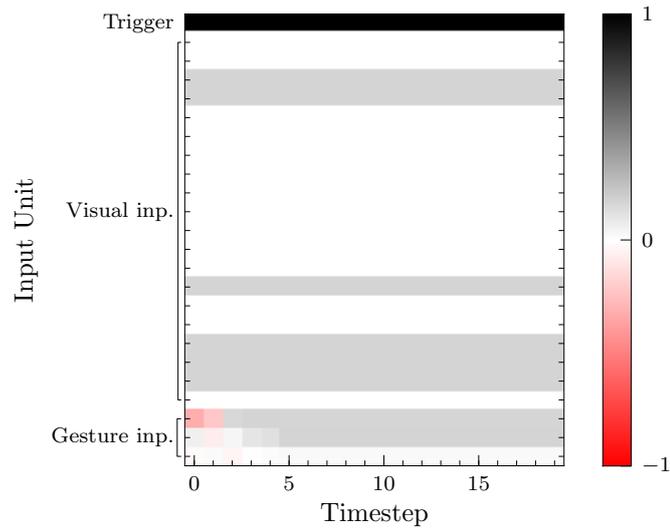


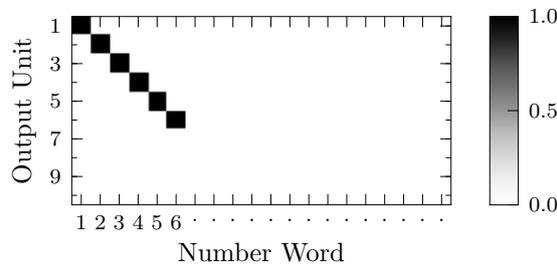
Figure 15: Example experimental design of a neuro-robotic simulation of counting. The figure refers to the simulation experiment described in section 6.3. Cf. figure 6.

In every experimental condition, the sequences of the number words (and, optionally, gestures) produced by the model in response to the test stimuli are recorded and their correctness is assessed based on the comparison with the corresponding target values. The *counting accuracy*, defined as the number of the sets from the test data set counted without any errors, is used as the principal index of the performance of the model, and serves as the dependent measure in the statistical analysis of the effects of the model parameters. Based on the test examples which have been counted incorrectly, the counting errors committed by the model are determined, and classified using the same criteria as those applied by Alibali and DiRusso (1999) to children counting (see figure 7 in chapter 4). Figures 16 and 17 illustrate the process of the network output evaluation in the correct and incorrect case, respectively.

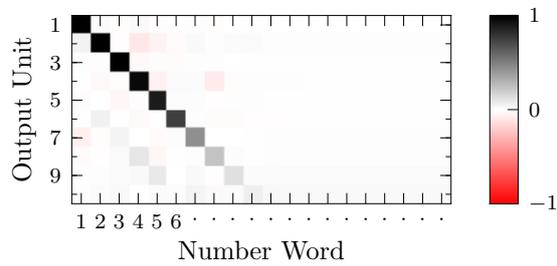
An additional measure of the model performance employed in some of the simulations is the *counting duration*, that is the length of the output produced by the model, defined as the last time step of the simulation sequence, at which the model utters a number word, and which is followed by silence until the end of the simulation. Note that although this does not take into account the correctness of the produced number words sequence, regressing the model counting duration against the size of the counted set (in other words, the actual length of the model output against the correct length) for all sequences in the test data set nevertheless provides useful insights into the behaviour of the model. The slope of the resulting regression line can serve as a marker of the quality of the behaviour of the model based on the following observations. For a model that counts perfectly correctly, the actual output length is always equal to the number of the counted items, and the resulting slope of the regression line is equal to 1 (and its intercept is 0). In turn, for a neural network that does not count, but simply produces a sequence of a fixed length regardless of the size of the counted set, the slope is equal to 0 (and the intercept indicates the length of the produced sequence). Intermediate values indicate a situation in-between and the value of the slope can be considered a quantitative indicator of ‘how hard the model is trying’ to count the items in its visual input.



(a)

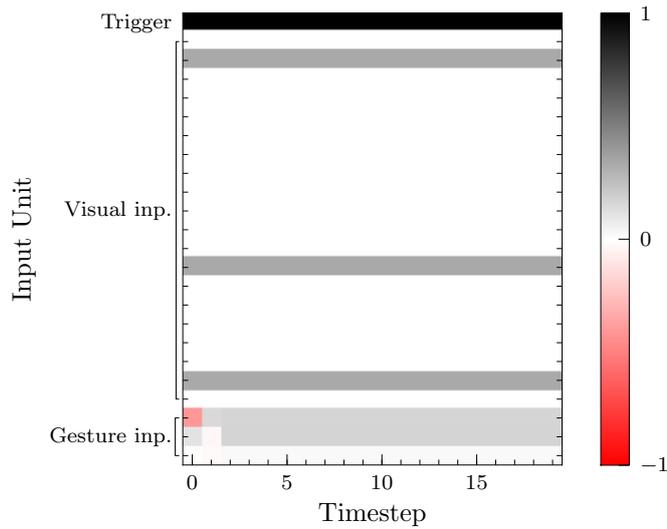


(b)

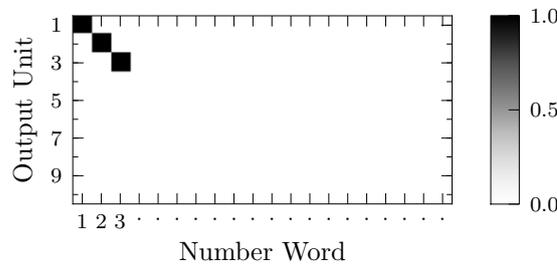


(c)

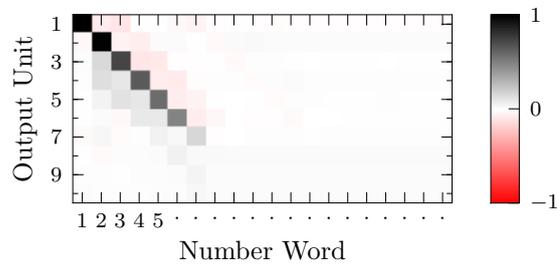
Figure 16: Example of a correct output of the model. The figure shows raster plots of the inputs to the neural network (a), the desired (i.e. target) network output (b), and the actual output produced by the network (c). In all three charts, the abscissa corresponds to the simulation time steps, the ordinate to the index of the unit in the appropriate layer of the neural network, and the intensity to the activation value of the unit. Below charts (b) and (c), a result of the nearest-neighbour classification of the network output at the given time step is shown, i.e. the number word assumed to be produced by the network (‘.’ denotes ‘silence’). In this example, there are six items in the visual input to be counted (in groups of 2, 1, and 3). Even though the actual network output (c) is not identical to the target output (b), in this trial, the network counting is considered to be correct. This is because, at every time step, the nearest-neighbour classification of the actual output is identical with that of the target output.



(a)



(b)



(c)

Figure 17: Example of an incorrect output of the model. The figure shows raster plots of the inputs to the neural network (a), the desired (i.e. target) network output (b), and the actual output produced by the network (c). In all three charts, the abscissa corresponds to the simulation time steps, the ordinate to the index of the unit in the appropriate layer of the neural network, and the intensity to the activation value of the unit. Below charts (b) and (c), a result of the nearest-neighbour classification of the network output at the given time step is shown, i.e. the number word assumed to be produced by the network (‘.’ denotes ‘silence’). In this example, there are three items in the visual input to be counted. Note that the network continues to count up to five, beyond the number of the items in the set; therefore, in this trial, the network counting is not considered to be correct, and the network commits the ‘Continue’ error.

5.5 Model Implementation

The proposed neural network model has been implemented using the PyBrain artificial neural network library for Python (Schaul et al., 2010). This significantly contributed to the portability of the implemented code. All inputs and outputs to the neural network were implemented using units with a linear activation function. Units of the hidden layer used the logistic activation function. The training algorithms (RProp- and backpropagation through time) were used in the implementations provided by the PyBrain library.

5.6 Discussion

As the means of summarising the description of the model presented in this chapter, it is appropriate to discuss the proposed architecture in the light of the past efforts to model various aspects of counting, reviewed in chapter 3.

In contrast to the models of Amit (1988), Hoekstra (1992), and Rodriguez et al. (1999), the model counts static, rather than sequential stimuli. It can be argued that this corresponds more closely to the context in which the children gain the majority of their initial experience with counting. When the children enumerate toys they are playing with, or pictures in a book they read with their parents, the set being counted is static, and its numerosity is related with its spatial, rather than temporal, characteristics. Whereas counting sequential (e.g. auditory) stimuli is of course a realistic scenario, it is not the one that primarily contributes to the children's acquisition of the counting skill. The distinction between enumerating spatially and temporally conveyed sets is particularly important in the context of the investigation of the contribution of the gestures to learning to count (what is tightly connected with the research questions 2 and 3 which the model aims to address). As discussed earlier in chapter 2, the establishment of a correspondence between the spatial aspect of the visually presented set and the temporal aspect of the recited count list is one of the crucial elements of mastering counting, and the

one in which the gestures are likely to be particularly helpful. It should be clear therefore, that in the present study it is most appropriate to consider static stimuli.

Although learning to recite the sequence of number words is one of the important aspects of the training regime of the proposed model (section 5.3.1), it is worth to stress that a detailed reproduction of the subtleties of the equivalent process in children is not the main focus of the present work. Ma and Hirai (1989) proposed a model which is capable of explaining many phenomena that appear in this context. While a simulation which focuses on the learning of the counting list is conducted in chapter 6, the proposed neural network is not expected to exhibit all the minute details of this process, such as those connected with the phonetic similarity between certain number words in the English language (Fuson et al., 1982). This is because the mechanism of learning sequences in the employed neural network architecture is likely not the best available model of rote learning in humans. More emphasis in the simulations will be put on how mastering the count list may affect the subsequent process of learning to count (in connection with the research question 1).

Many similarities can be found between the model described herein and the one proposed by Ahmad et al. (2002), however there are important differences between these approaches. A prevailing theme in the models of Ahmad et al. is the application of the mixture of experts architecture on several levels of the model design. While it is no doubt an elegant and interesting machine learning device, Ahmad et al. provide no justification, e.g. in form of the evidence of its biological plausibility, for employing this solution so abundantly. Furthermore, at least in some instances where Ahmad et al. insert the mixture of experts model, it can be argued that its application is superfluous. An example worth focusing on is the counting module of their SCOUSYST model (Ahmad et al., 2002, pp. 187–197). It is composed of two subsystems, *word*, a simple recurrent network (Jordan type), and *next-object*, a feed-forward network. Ahmad et al. argue that the mixture of experts architecture on top of those ‘selects an expert network that is optimal for the designated subtask’ (Ahmad et al., 2002, p. 189), delegating each of the subtasks (production of words

and of indicating acts) to the appropriate sub-network. However, the fact that a Jordan network is an extension of a feed-forward network, and is therefore perfectly capable of learning the exact same mappings as the latter, suggests that such a solution is complicated above what is necessary. What makes matters worse, since the gestures and the recitation of the number words are implemented by separate neural networks, the internal representations employed in those tasks are prevented, *by the model design*, from interacting. This is in stark contrast with the behavioural data reviewed in chapter 2, which indicate that, during learning to count, the gestures affect the counting performance substantially. The situation just described is a prime example of the problem discussed while introducing the design assumption 3 at the beginning of this chapter. While the architecture of Ahmad et al. is no doubt capable of exhibiting a counting-like behaviour, the fact it is unnecessarily overcomplicated limits its usability as an aid in the cognitive study of counting. The design of the model of counting proposed in this thesis is significantly simpler than that of Ahmad et al., as one of the design objectives was to endow the model with the capabilities necessary to perform the task, but not to bias its operation without sufficient justification.

Finally, it is worth to highlight an important feature of the design of the proposed model of learning to count, namely its considerable flexibility, which allows a wide array of modelling scenarios to be tested in a unified way. This flexibility has at least two dimensions. The first one is connected with the set-up of the neural network — more specifically, the counting gestures may either be an input to or an output from the model (cf. section 5.2.2). The second one results from the fact that the model is compatible with a variety of representations. The latter means that the proposed neural network may be used to compare the consequences of employing different approaches to represent information in the context of counting. This applies equally to the representation of the speech, counting gestures, as well as the visual information. Since several different approaches are possible for each modality, the investigation of all realisable scenarios was unfortunately not possible within the

scope of the present thesis. Thus, it is important to keep in mind that the utility of the proposed model goes beyond addressing only a few research questions considered herein.

Summarising, the model presented in this chapter constitutes a novel contribution to the state-of-the-art in modelling the role of gestures in learning to count in that:

- this is the first cognitive model in the context of mathematical cognition designed according to the developmental cognitive robotics paradigm (cf. section 4.2), closely linked with an artificial body to represent embodied phenomena. Considering the ample evidence for the embodied nature of human numerical knowledge in general, and counting in particular (cf. chapter 2), this is an important step forward;
- as the consequence of the above, this is the first model to employ a realistic representation of the counting gestures based on the actual pointing performed by a humanoid robot endowed with dexterous arms that have been designed to resemble human arms as close as possible (cf. section 4.3);
- it is the first model of the *contribution* of the counting gestures to learning to count. Although one of the previously published models of counting incorporated gestures (Ahmad et al., 2002), several design decisions in that study severely biased the ways in which the gestures could affect the counting process. One of the aims in the present work is to minimise such biases;
- the proposed model is conceptually simple, but at the same time flexible with respect to the representations of the various aspects of counting (such as the encoding of the visual, proprioceptive and verbal information), what allows a wide variety of hypotheses to be tested using the model;

The proposed model is investigated in a series of simulations which are described, and the results of which are reported in the subsequent chapter.

Chapter 6

Simulations of Learning to Count using the Robotic Model

In this chapter the results of the simulations conducted using the model of learning to count introduced in chapter 5 will be discussed. The first two simulations have a preliminary character. The first one focuses on the initial stage of the training of the model, which corresponds to the learning of the count list. The second simulation investigates the ability of the model to generalise, that is to count sequences that have not been presented to the model during training, as well as looks at the influence of including the initial training phase on further training of the model, thus tackling the research question 1. Finally, the third and fourth simulation use the proposed model in experiments which investigate the contribution of the counting gestures to learning to count in attempts to answer research questions 2 and 3, respectively.

6.1 Simulation 1 — Learning Number Words

6.1.1 Aims of the Experiment

The aim of this simulation was to gain knowledge about the progress of the first stage of the training of the model. As explained in section 5.3.1, this preliminary stage has been introduced to reflect the ontogeny of counting in children, who are able to recite the count list quite well before they start learning to count things. In

line with this, before the model is trained to count using vision and gestures, it is equipped with the ability to produce a sequence of number words, as if learnt by rote. This simulation provided certain basic information about the model and its training that were useful in determining the values of the parameters for subsequent simulations, such as the number of hidden units in the network or the length of the training. Although, as explained in chapter 5, the detailed modelling of the learning of the count list has not been among the main goals of the model, it was nevertheless interesting to compare the behaviour of the model with human data.

6.1.2 Procedure

In this simulation experiment only the first stage of the training was performed. The model was therefore used in its initial configuration, which does not include any of the optional modules (see figure 12). It was assumed that the model is trained to recite $N_{VO} = 10$ number words. A generous amount of training was given to the model (10000 epochs) in order to assure that the learning process has enough time to converge. The other training parameters were as reported in section 5.3.1.

The number of hidden units in the network N_H was the main parameter that was varied in order to investigate how it affects the results of the training. Based on earlier informal trial-and-error experiments with the model in various configurations, the following values of N_H have been chosen for systematic investigation: 6–12, and 15, 20, and 25. The 6–12 range has been estimated to contain the minimum number of units required for the network to successfully acquire the preliminary skill. The simulation aimed thus to determine the exact value of this limit. The latter three values were the numbers of hidden units chosen for the subsequent simulations, which include the second stage of the model training. These were considered in the present simulation to establish the required duration of the preliminary training.

After the training of the model was completed, the neural network was evaluated in order to determine if the training has been successful, using the criteria described in section 5.3.1. In addition, during the training the information about the learning

progress were collected. After every training epoch, the mean-squared error over the training data set, and the sequence of number words produced by the model in response to the trigger input were recorded. If the produced count list changed with respect to the previously recorded one, the number of the epoch at which the change occurred was recorded. Using these information it is possible to determine the epoch of the training at which each number word has been ‘acquired’ by the network, as well as the number of the training epochs necessary for the model to master the complete count list.

For each of the considered sizes of the hidden layer of the neural network 10 independent repetitions of the training were performed, with different initial weights of the connections, in order to estimate the training success rate for each hidden layer size. Since 10 distinct values for N_H were considered, this resulted in the total of 100 repetitions of the simulation.

6.1.3 Results

The number of trials in which the training was successful for each of the considered sizes of the hidden layer N_H are reported in table 1.

As a means of illustrating a typical progress of the number word learning in the proposed model in various cases, the evolution of the network output throughout the first stage of training for three experiment trials is reported in figures 18, 19, and 20. These figures show the number of the epoch of the training in which the output sequence first appeared and the output sequence itself. Numbers 1–10 designate the corresponding number words, while the dots represent silence (all output units of the network deactivated). The three instances of the training were chosen as representing typical model behaviour. Figure 18 represents an unsuccessful training case. The training in the trials illustrated in figures 19 and 20 was successful, but the latter progressed faster than the former, due to the larger size of the hidden layer.

Epoch	Model Output																
1
2	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
3
4	8
5
73	1
74
75	1
76
79	1
260	1	2
261	1
262	1	2
264	1
265	1	2
608	1	2	3
610	1	2
611	1	2	3
656	1	2	3	.	.	6
659	1	2	3	4	.	6
660	1	2	3	.	.	6
661	1	2	3	4	.	6
662	1	2	3	.	.	6
665	1	2	3	4	.	6
666	1	2	3	4
667	1	2	3	.	.	6
668	1	2	3	4	.	6
952	1	2	3	4	.	6	7
983	1	2	3	4	5	6	7
984	1	2	3	4	.	6	7
995	1	2	3	4	5	6	7
1320	1	2	3	4	5	6	7	8
1321	1	2	3	4	5	6	7
1325	1	2	3	4	5	6	7	8
1326	1	2	3	4	5	6	7
1327	1	2	3	4	5	6	7	8
1332	1	2	3	4	5	6	7
1333	1	2	3	4	5	6	7	8
1334	1	2	3	4	5	6	7
1335	1	2	3	4	5	6	7	8
1336	1	2	3	4	5	6	7
1337	1	2	3	4	5	6	7	8
1338	1	2	3	4	5	6	7
1339	1	2	3	4	5	6	7	8
1683	1	2	3	4	5	6	7	8	9
1692	1	2	3	4	5	6	7	8
1693	1	2	3	4	5	6	7	8	9
1699	1	2	3	4	5	6	7	8
1700	1	2	3	4	5	6	7	8	9
6090	1	2	3	4	5	6	7	8
6091	1	2	3	4	5	6	7	8	9
6106	1	2	3	4	5	6	7	8
6107	1	2	3	4	5	6	7	8	9
6108	1	2	3	4	5	6	7	8
6109	1	2	3	4	5	6	7	8	9
6110	1	2	3	4	5	6	7	8
6130	1	2	3	4	5	6	7	8	9
6131	1	2	3	4	5	6	7	8

Figure 18: Number words learning progress in trial 028 ($N_H = 8$, training not successful)

Increasing the size of the hidden layer affected the speed with which the model acquired the number words. Figure 21 shows the distributions of the numbers of the training epoch from which on the network began to use the number words correctly across trials with 15, 20 and 25 hidden units.

6.1.4 Discussion

The success rates reported in table 1 indicate that at least 11 units in the hidden layer of the proposed neural network model are necessary in order for it to be able to reliably learn a sequence of 10 number words. Although some successful attempts appear for N_H as small as 9, low success rate suggests that the training algorithm in this case easily gets stuck in the local optima of the training error landscape. These results are not surprising considering the employed type of the output layer (i.e. linear), which imposes that the states of the hidden layer units have an analogous topology as the output vectors they correspond to. Since all output vectors that represent number words are orthogonal to each other (cf. section 5.2.3), the same has to be true for their corresponding internal representations — and since in the simulation $N_{VO} = 10$, this requires the internal representation space to be at least 10-dimensional. Occasional appearance of the successful training attempts in the 9-dimensional case can be explained by the fact that the employed protocol of the model evaluation, based on nearest-neighbour classification, allows for a degree of imprecision at the network output. The 10 internal states in those successful cases must have been packed into the 9-dimensional space in such a way that, although not exactly orthogonal to each other, they were close enough to being in such a state.

There are several characteristic features of the progress of the number word learning in the proposed model that can be observed in figures 18, 19, and 20. Initially, the output of the model is a random number word that usually does not change over time. This is the consequence of the fact that before the training the weights of the neural network are initialised randomly. After a few training epochs,

Epoch	Model Output																		
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	7	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
3
34	1
60	1	2
115	1	2	3
116	1	2
117	1	2	3
118	1	2
124	1	2	3
125	1	2
126	1	2	3
127	1	2
128	1	2	3
189	1	2	3	4
190	1	2	3
191	1	2	3	4
276	1	2	3	4	5
435	1	2	3	4	5	6
436	1	2	3	4	5
437	1	2	3	4	5	6
527	1	2	3	4	5	6	7
578	1	2	3	4	5	6	7	8
579	1	2	3	4	5	6	7
624	1	2	3	4	5	6	7	8
762	1	2	3	4	5	6	7	8	9
763	1	2	3	4	5	6	7	8
764	1	2	3	4	5	6	7	8	9
1310	1	2	3	4	5	6	7	8
1311	1	2	3	4	5	6	7	8	9
1313	1	2	3	4	5	6	7	8
1318	1	2	3	4	5	6	7	8	9
1319	1	2	3	4	5	6	7	8
1712	1	2	3	4	5	6	7	8	9
1714	1	2	3	4	5	6	7	8
1715	1	2	3	4	5	6	7	8	9
1717	1	2	3	4	5	6	7	8
1718	1	2	3	4	5	6	7	8	9
1750	1	2	3	4	5	6	7	8
1751	1	2	3	4	5	6	7	8	9
1752	1	2	3	4	5	6	7	8
1753	1	2	3	4	5	6	7	8	9
1758	1	2	3	4	5	6	7	8
1759	1	2	3	4	5	6	7	8	9
1782	1	2	3	4	5	6	7	8
1783	1	2	3	4	5	6	7	8	9
1788	1	2	3	4	5	6	7	8
1789	1	2	3	4	5	6	7	8	9
1793	1	2	3	4	5	6	7	8
1794	1	2	3	4	5	6	7	8	9
1799	1	2	3	4	5	6	7	8
1800	1	2	3	4	5	6	7	8	9
1805	1	2	3	4	5	6	7	8
1806	1	2	3	4	5	6	7	8	9
1812	1	2	3	4	5	6	7	8
1813	1	2	3	4	5	6	7	8	9
2867	1	2	3	4	5	6	7	8	9	10
2868	1	2	3	4	5	6	7	8	9
2910	1	2	3	4	5	6	7	8	9	10
2911	1	2	3	4	5	6	7	8	9
2916	1	2	3	4	5	6	7	8	9	10
2917	1	2	3	4	5	6	7	8	9
2918	1	2	3	4	5	6	7	8	9	10
2919	1	2	3	4	5	6	7	8	9
2920	1	2	3	4	5	6	7	8	9	10

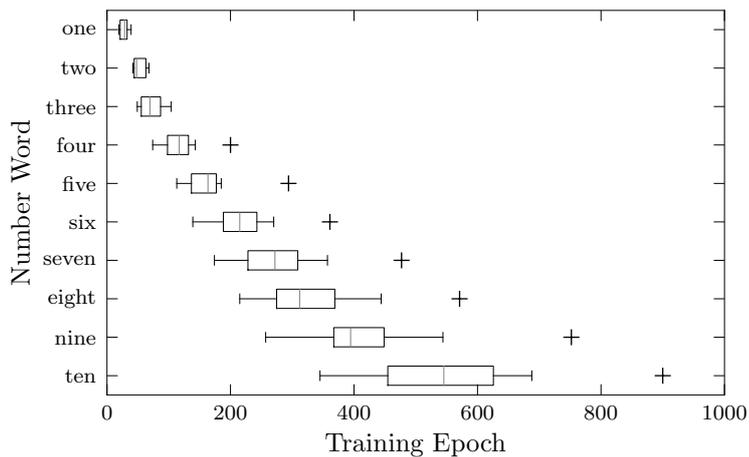
Figure 19: Number words learning progress in trial 042 ($N_H = 10$, training successful)

N_H	6	7	8	9	10	11	12	15	20	25
Successful Trials	0	0	0	3	8	10	10	10	10	10

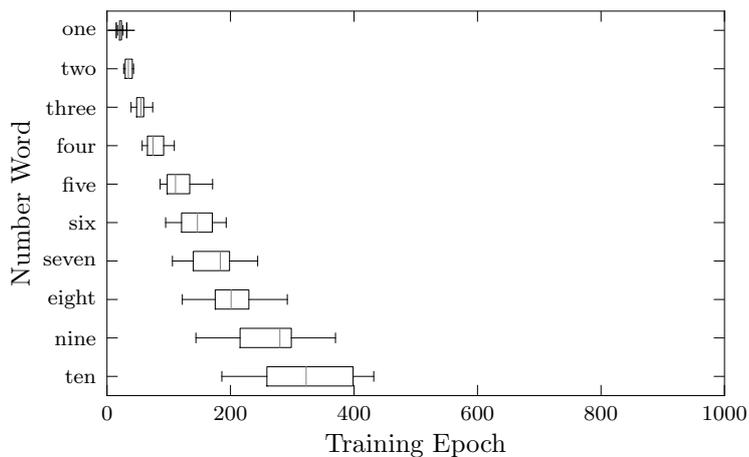
Table 1: Number of trials (out of 10) in which the preliminary training stage was successful for the considered hidden layer sizes N_H

Epoch	Model Output																
1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2	7	.	7
3
4	3
5
31	1
64	1	2
73	1	2	3
132	1	2	3	4
185	1	2	3	4	5
244	1	2	3	4	5	6
309	1	2	3	4	5	6	7
333	1	2	3	4	5	6	7	8
363	1	2	3	4	5	6	7	8	9
364	1	2	3	4	5	6	7	8
365	1	2	3	4	5	6	7	8	9
367	1	2	3	4	5	6	7	8
368	1	2	3	4	5	6	7	8	9
612	1	2	3	4	5	6	7	8	9	10
613	1	2	3	4	5	6	7	8	9
627	1	2	3	4	5	6	7	8	9	10

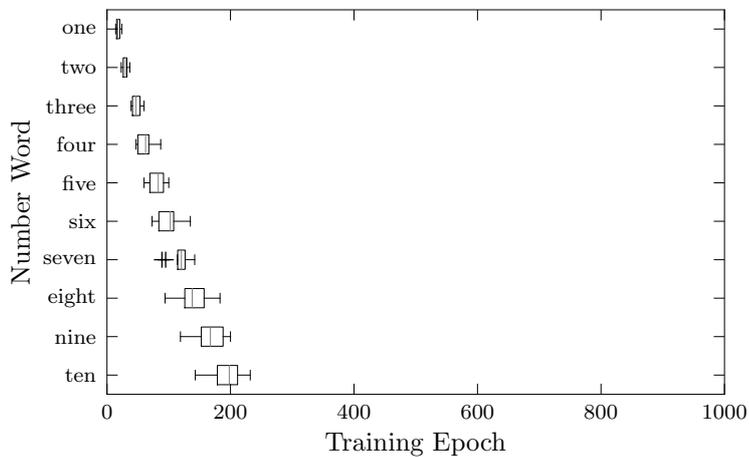
Figure 20: Number words learning progress in trial 075 ($N_H = 15$, training successful)



(a) $N_H = 15$



(b) $N_H = 20$



(c) $N_H = 25$

Figure 21: Progress of number words learning across 10 trials for the neural networks with 15 (a), 20 (b), and 25 (c) units in the hidden layer. This is a box plot that illustrates the distributions, across the 10 trials, of the epoch numbers, in which the number words were acquired by the model. The boxes extend from the lower to the upper quartile of the data, the whiskers extend to the most extreme data points (limited to 1.5 times the quartile range), the grey line indicates the median, and outliers are shown using the + symbol.

this random output is suppressed and the model begins to learn the correct count list, starting from the beginning. Normally the learning progresses with consecutive number words, but occasionally a number word can be acquired by the model outside of the conventional sequence (e.g. number word ‘six’ in figure 18, epoch 656). The latter phenomenon did not appear frequently and tended to occur for the lower values of N_H . One of the most distinct features of the learning are the prolonged phases in which a number word is interchangeably produced by the model and not, before it starts to appear at the output consistently (e.g. in figure 19 this phenomenon is evident for number words ‘three’, ‘four’, ‘six’, ‘eight’, ‘nine’, and ‘ten’). As more hidden units are added to the model, these phases are visibly shortened, but can still be observed even for N_H larger than the required minimum (cf. figure 20). It is important to note that even for the small sizes of the hidden layer and the cases of unsuccessful training, the model did not produce the number words outside of their corresponding time steps, except for the very early stages of the training. In the cases of unsuccessful training the model did not manage to learn all 10 number words, but only a certain amount of the beginning of the count list (see e.g. figure 18).

The progress of the learning of the count list by the proposed model resembles to a limited degree the equivalent process in children (cf. section 2.3.2). Since, once trained, the employed neural network framework is deterministic (i.e. the output produced by the model is always the same for the same input to the network), the stochastic within-subject effects present in the children’s behaviour, such as the unstable non-conventional portion of the count list, are not reproduced. As illustrated in figures 18, 19, and 20, the model also does not seem to have a tendency to emit stable non-conventional sequences. Rather, at each time step of the simulation the model either produces a correct number word, or does not utter anything (an exception to this, as mentioned above, are the very early epochs of the training). The behaviour of the model is however consistent with that of children with respect to the acquisition of the stable conventional portion of the count list. The count list

produced by the model starts with small numbers and is gradually extended toward the larger ones. Furthermore, prolonged phases during which a consolidation of the most recent extension takes place can be clearly distinguished (see e.g. figure 19, cf. Fuson et al., 1982).

Based on figure 21 it is possible to determine the required duration of the preliminary training stage for subsequent simulations. Since in the further experiments models with $N_H = 15, 20$ and 25 are used (with N_H acting as a between-subjects factor), the number of the epochs in the first stage of the training should be sufficient to ensure reliable success of this stage for all considered values of N_H . As figure 21 indicates, the amount of training required for the model to learn to recite 10 number words is the largest for $N_H = 15$. Since in 9 out of 10 cases 700 epochs were sufficient for the model to master the entire count list (there was only one outlier case for which around 900 epochs were necessary — see figure 21a), this number has been used as the duration of the preliminary stage in all subsequent simulations.

6.2 Simulation 2 — Impact of the Preliminary Training Stage and Generalisation

6.2.1 Aims of the Experiment

Inclusion of the preliminary training stage in the training regime of the proposed model is intended to resemble how learning to count progresses in children. Since such an approach is not standard from the point of view of the usual practice of neural network training, it is appropriate to examine if this has any impact on the final performance of the proposed model. At the same time, such an experiment is expected to provide an answer to the first research question posed in the present thesis: how does mastering the count list prior to learning to count within the respective range of collection sizes affect the subsequent process of learning to count? Addressing this question was the primary aim of this simulation. In addition, the

ability of the model to generalise, that is to count arrangements of items that the model has never encountered during training, was assessed.

6.2.2 Procedure

In this simulation the model was used in the configuration depicted in figure 22, that is it was trained to count using both visual and proprioceptive information, with the latter acting as an input to the network. This configuration has been chosen as this is the primary model set-up used in the later simulations. In order to investigate the impact of the preliminary training stage on the outcome of the second training stage, the training protocol introduced in section 5.3 was modified in the following way. Once the model in the initial configuration was created (figure 12), a copy of it was made, in order to preserve the identical initial weights of the network connections. One of the copies was then subjected to the preliminary training stage followed by the second stage, while the other proceeded directly to the second stage of the training. The first network thus acquired the ability to recite a sufficiently long count list before proceeding with learning to count items, while the second one had to acquire both these skills at the same time. In both cases, during the second stage of the training identical training data sets were used (the training data sets were of course different between trials, as were the initial weights in the network). In order to compensate for the additional training received by the first network, the second stage of the training of the second network was prolonged from 4000 to 4064 epochs (there are 2 sequences in the training data set in stage 1 and 22 in stage 2, and therefore 700 training epochs in stage 1 correspond, in terms of the number of performed weights updates, to approximately 64 epochs in stage 2). The training parameters in both stages were as reported in section 5.3.

The experiment was repeated 30 times for $N_H = 15, 20$ and 25 , yielding the total of 90 trials. After training, the models were evaluated on two types of test data sets. The first data set, identical for all 90 trials, consisted of 50 collections (5 different examples for every number from 1 to 10) in arrangements that have

never been shown to any of the networks during the training. The second data set, specific for every trial (but identical for the two networks within a trial), consisted of 50 arrangements of items chosen from those shown to the networks during training. The number of examples from a test data set counted correctly served as an index of the performance of the model.

6.2.3 Results

The experimental set-up described above corresponds to a mixed-design $2 \times 2 \times 3$ (stage 1 presence or absence \times test data set known or unknown \times $N_H = 15, 20,$ or 25) repeated-measures ANOVA, with N_H as the between-subjects factor and the number of test examples counted correctly as the dependent measure. The difference in counting accuracy caused by inclusion versus omission of the preliminary training stage and by the type of the test data set were two planned contrasts. In 3 trials, all with $N_H = 15$, the preliminary training stage did not finish with a successful acquisition of the count list by the model. These trials were therefore discarded, what left 87 trials available for the statistical analysis. Statistically significant effects of the stage 1 inclusion ($F = 965.954, p < 0.001, \eta_p^2 = 0.920$), of the training data set type ($F = 105.278, p < 0.001, \eta_p^2 = 0.556$), and that of N_H ($F = 6.739, p = 0.002, \eta_p^2 = 0.138$) were found. In addition, the interaction between the within-subject factors was significant ($F = 10.713, p = 0.002, \eta_p^2 = 0.113$). N_H did not interact with the within-subject factors. The profile plots of the estimated marginal means for the stage 1 versus data set type interaction and for N_H are shown in figure 23.

The effect of N_H as well as the interaction between the within-subjects factors were investigated in post-hoc analysis, with assumed level of significance $\alpha = 0.01$. Pairwise comparisons between the three levels of N_H (with α adjusted for multiple comparisons using the Holm-Bonferroni method) indicated the only statistically significant difference to be $N_H = 15$ against $N_H = 25$ ($p = 0.002$, the other p values were $p = 0.035$ for $N_H = 20$ against $N_H = 25$ and $p = 0.924$ for $N_H = 15$ against $N_H = 20$). Pairwise comparisons for the stage 1 inclusion versus dataset

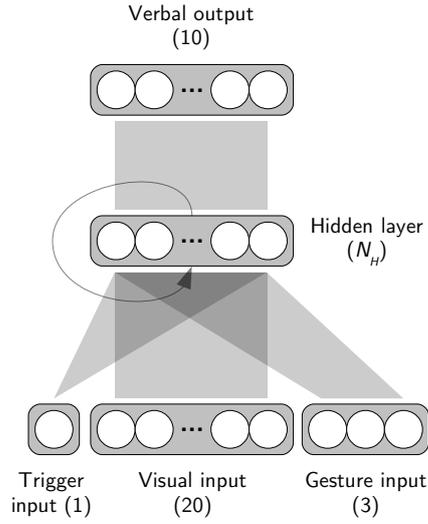


Figure 22: Configuration of the model used in simulation 2. $N_{VO} = 10$, $N_{VI} = 20$, and $N_G = 3$.

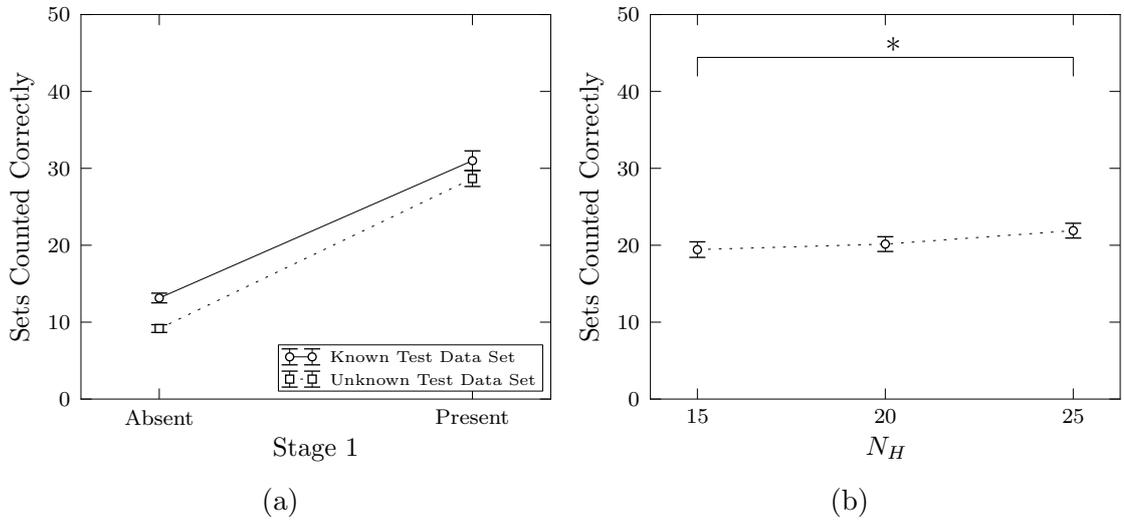


Figure 23: Profile plots for the simulation 2 ANOVA. The plots show the number of examples from the test data set (out of 50) counted correctly by the model after the second stage of the training. (a) illustrates the stage 1 inclusion versus dataset type interaction. (b) shows the effect of the size of the hidden layer in the model. The error bars indicate 95% confidence intervals. All possible pairwise comparisons in (a) are significant at $\alpha = 0.01$ (see text). The star indicates the only statistically significant difference (at $\alpha = 0.01$) in (b).

type interaction, based on a paired-samples t -tests adjusted for multiple comparisons using the Holm-Bonferroni method, indicated all 6 comparisons to be statistically significant at assumed α (all $p < 1 \cdot 10^{-5}$). As evident in figure 23a, the discovered interaction is of the ordinal type, therefore its presence does not invalidate the discussion of the discovered main effects of stage 1 inclusion and of the dataset type.

6.2.4 Discussion

The statistical analysis indicates that the counting accuracy of the model was affected by three factors:

- whether the model was tested on known or unknown item arrangements;
- whether the preliminary training stage was included or not;
- the number of hidden units in the network.

As shown in figure 23a, the proposed neural network tended to achieve better scores on the known test data set than on the unknown one, and when the preliminary training stage was included than when it was not. The latter is an especially important result in the context of the considered research question. It shows that, in addition to being justified from the theoretical standpoint by the data from experimental psychology, the introduction of the preliminary training stage brought tangible benefits in terms of the final counting performance of the model. Note that, in addition, a statistically significant interaction between the stage 1 inclusion and the dataset type has been found. The slopes in figure 23a indicate that when the preliminary training stage was included in the model training, the counting performance of the model was less affected by whether the model has been tested on the known or on the unknown arrangements of items. In other words, the preliminary training stage not only allowed the model to achieve higher counting accuracy within the given amount of training, but also enabled it to generalise better.

The mechanism explaining the contribution of the preliminary training stage on the model's counting accuracy is most likely the following. Successful preliminary

training stage equips the model with an ability to produce a correct count list, the length of which is sufficient to count any set that is presented to the model during the second training stage. The task of the model during the second stage is therefore only to learn to ‘modulate’ its output based on the visual and proprioceptive information present at input, rather than to learn the entire task from scratch. Evidently, the state in which the weights of the network are left after the preliminary stage biases the subsequent training in such a way that better counting accuracy can be achieved within the comparable amount of training. Although limited to the scope of the considered largely simplified scenario, the above finding may be interpreted as computational evidence that being equipped with a sufficiently long count list prior to learning to count collections within the respective number range makes the latter task easier. While this does not mean that children’s acquisition of the appropriate portion of the count list before learning to use it to count items is a necessary developmental step in the Piagetian sense, it is possible that this allows the subsequent learning process to be sped up, perhaps to the point where it actually happens within a reasonable time, or even within the lifetime of the individual.

As the conducted statistical analysis indicates, the network’s final counting performance was also affected by whether the test data set consisted of the item arrangements which were shown to the network during the training, or not. However, the size of this effect was marginal in comparison to that of the preliminary training stage, as seen in figure 23a and through the obtained η_p^2 values. Overall, it can be said that the model generalised quite well from the arrangements on which it has been trained to the novel sets of items, with only a modest drop in the counting accuracy. In accordance with this finding, in the subsequent simulations the test data sets were composed solely of item arrangements that have not been shown to the model during the training.

Finally, the counting accuracy was also affected by N_H , although to a rather modest degree (as indicated in figure 23b and by the low η_p^2). Importantly, the size of the hidden layer in the network did not interact with the other factors, which

means that the discovered effects were robust across the considered range of N_H . Since the size of the hidden layer did not affect, in the considered range of its values, the ability of the proposed neural network to generalise, in subsequent simulations N_H was simply fixed to 20.

6.3 Simulation 3 — Contribution of the Counting Gestures to Learning to Count

6.3.1 Aims of the Experiment

The main purpose of the developmental neuro-robotic model introduced in chapter 5 is the investigation of the contribution of the counting gestures to learning to count. This goal is achieved in this simulation by assessing the counting performance of the model across different experimental conditions, looking for the evidence that the pointing gestures have improved the counting accuracy. The primary aim of this experiment is to seek quantitative evidence for the usefulness of the proprioceptive information connected with the gestures in the context of the counting task, beyond known behavioural studies, and therefore providing an answer to the research question 2. In addition, the simulations are expected to provide additional insights into the exact nature of this contribution.

6.3.2 Procedure

As explained in chapter 5, section 5.4, the experimental set-up employed in the present simulation was largely inspired by the one designed by Alibali and DiRusso (1999) to study the function of the counting gestures in children. The ‘subjects’ of the experiment were tested in multiple experimental conditions, in which the counting accuracy was used as the principal indicator of the performance. As in the quoted study, a distinction between small and large sets of items was made to investigate if the counting accuracy of the model depends on the size of the collec-

tion being enumerated. Also, the behaviour of the model was assessed qualitatively, which included the identification of the types of the counting errors and the frequencies with which they occurred across the conditions, using the same criteria as Alibali and DiRusso applied to children.

The present simulation exploited the flexibility of the proposed neural network by training it to count the items presented to its visual input in three experimental conditions. The corresponding model configurations are depicted in figure 24. Figures 24a and 24b represent counting without and with the counting gestures respectively, and are intended to resemble the equivalent conditions in the experiments with children. The only difference between these two conditions is that in the configuration shown in figure 24b, the proprioceptive information was available to the model as an input, in addition to the visual information. In the configuration shown in figure 24a, the model was trained to count based on the visual information only. Hence, a difference in the counting accuracy between these two configurations could confirm the importance of the proprioceptive information in the context of learning to count, and thus give the positive answer to the research question 2.

As evident in figure 24b, in this simulation it has been assumed that the proprioceptive information connected with the counting gestures are an external input to the model. In other words, it is not the task of the model to *produce* the counting gestures. As discussed in section 5.2.2, this is one of the viable alternatives. This particular option was selected, because both approaches are justified to some extent from the theoretical point of view, but the former represents a simpler model configuration. The choice is therefore consistent with the ‘information-theoretic’ stance of using a connectionist model as a learning-based mechanism of information processing, assumed in this thesis (cf. design assumption 1 in section 5.1).

The decision to provide the information connected with the gestures as an input to the neural network has at least two important consequences. First, it means that the simulated scenario corresponds more closely to the passive, rather than the active gesture conditions in the studies with children (see section 4.5). Second, in the model

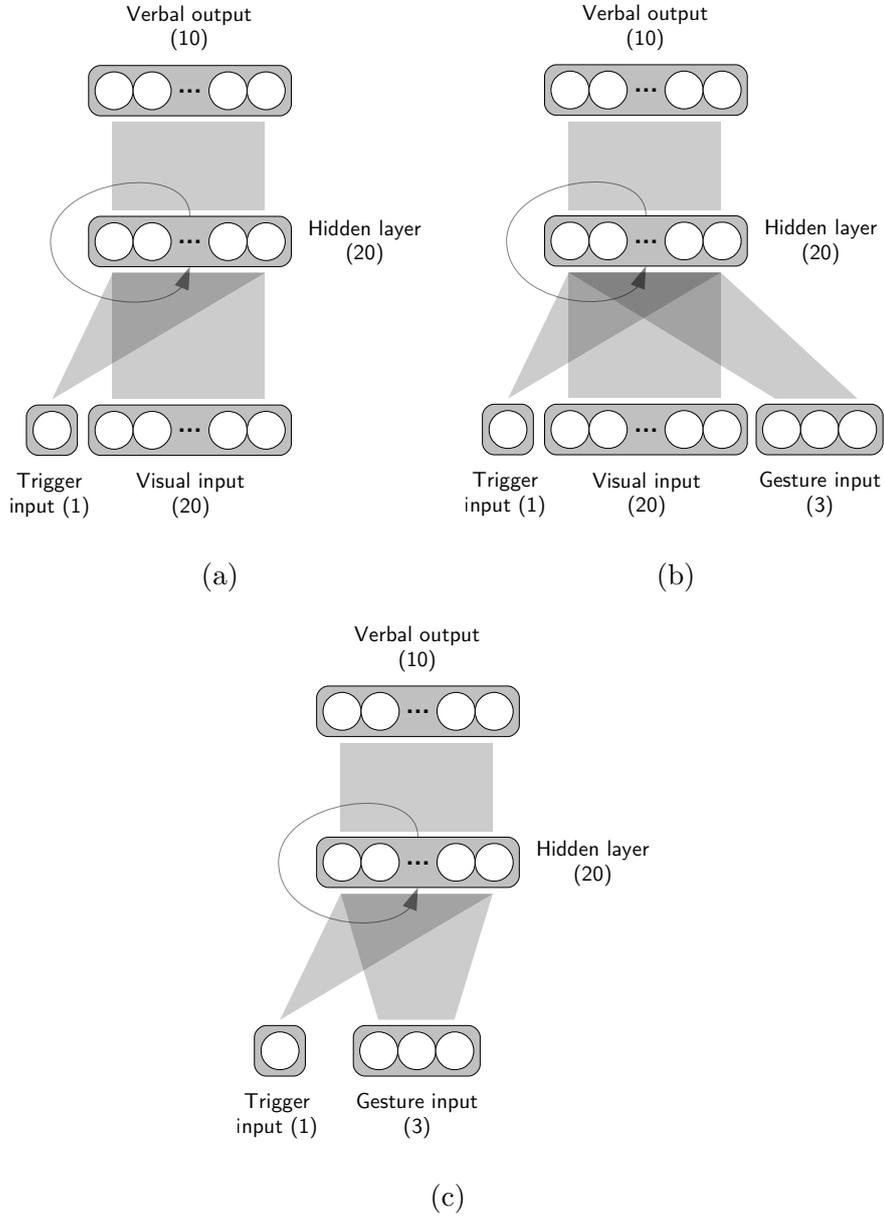


Figure 24: Configurations of the model used in simulation 3. Figure (a) shows the model in the V^+G^- (vision, no gestures) condition, figure (b) in the V^+G^+ (vision and gestures) condition, and figure (c) in the V^-G^+ (no vision, gestures) condition. In all conditions $N_{VO} = 10$, $N_{VI} = 20$, $N_G = 3$, and $N_H = 20$.

configurations where the gestures are provided, they must be *correct* from the point of view of counting, and, because of that, in the conducted simulations certain types of the counting errors considered by Alibali and DiRusso (1999) cannot appear. In particular, this is the case for the double count error (see figure 7). Note however that a similar situation occurred in the puppet conditions in the quoted behavioural study, where the gestures were performed by the experimenter and therefore they were also always correct.

The configuration of the model used in the third experimental condition, shown in figure 24c, corresponds to counting based on the proprioceptive input only, without the visual information available. This artificial condition, which is realisable in the proposed model but does not have a real-world equivalent in the studies with human participants, has been introduced in order to resolve a non-obvious ambiguity in the case of the (expected) positive outcome of the comparison between the other two conditions. Suppose, that the model in the configuration 24b indeed turns out to count better than the one shown in figure 24a. As explained earlier, since the only difference between these two set-ups is the presence of the proprioceptive input in the former, the increase in the counting accuracy is due to this additional information. It is not clear however, what use does to model make of the visual information after it is supplied with the proprioceptive input. It is possible that the improvement of the model's performance is solely due to the new information, and the former is actually completely disregarded. In order to find out if this is the case, the performance of the model that counts based on both vision and gestures needs to be contrasted with that based only on the gestures. Should the model achieve comparable counting performance in these two conditions, this would mean that in the former case the visual information is, most likely, ignored.

For convenience, from now on the three experimental conditions in this simulation will be referred to using the following self-explanatory abbreviations: vision, no gestures: V^+G^- , both vision and gestures: V^+G^+ , no vision, gestures: V^-G^+ .

The training protocol employed in the present simulation followed the one de-

scribed in section 5.3. The model was first trained to produce a sequence of ten number words, and then extended to yield the three experimental conditions. Across the conditions within one trial, the training data sets used in the second stage of the training were constructed based on the identical arrangements of items, presented in the same sequence. After the second stage of the training in all three experimental conditions was finished, the models were evaluated on a test data set, which consisted of the arrangements of items that have not been shown to any of the networks during the training. Based on the results of the earlier simulations, the value of the N_H parameter was fixed to 20. The above procedure was repeated independently 32 times, with the training data sets and the initial weights of the connections in the network randomised.

The evaluation protocol in the present simulation followed closely the one applied by Alibali and DiRusso (1999) to children’s counting (see section 5.4). The analysis of the results started with the statistical investigation of the factors influencing the counting accuracy of the model. Two planned contrasts compared the counting accuracy between V^+G^+ and V^+G^- conditions, and between V^+G^+ and V^-G^+ . Similarly to the quoted experimental study, the effect of the size of the counted set was also investigated. The examples in the test dataset were divided into *small numbers* (1–5) and *large numbers* (6–10), yielding a two-level factor, which was the subject of another planned contrast.

In addition to the statistical analysis described above, the behaviour of the model was also investigated qualitatively. All unsuccessful counting attempts over the test data set were assessed in terms of the presence of the counting errors made by the model. The same classification of the counting errors as used by Alibali and DiRusso (1999) was used for that purpose (see figure 7 in section 4.5).

Finally, the regression analysis of the actual length of the output of the model against the target length (see section 5.4) across the conditions was performed. A danger, which exists when training any artificial neural network is, that the learning process may get stuck in so-called local optimum. In the context of the proposed

model this could mean converging toward a behaviour, which yields a relatively small error on the training dataset, but which does not correspond to counting objects shown in the visual input. One particular behaviour which was feared to be likely to occur was the one where the network produces a sequence of number words corresponding to counting to 5 or 6 (which is approximately half of the largest considered number), regardless of the actual number of the objects in the counted set. Such a behaviour could constitute a local minimum in the training error landscape, because the average number of wrong number words produced per set would be minimised, and producing such a behaviour would not require much ‘cognitive effort’ from the network. As explained in section 5.4, by regressing the actual length of the model output against the size of the enumerated collection, it is possible to detect such deteriorated behaviour.

6.3.3 Results

The experimental set-up yields a 3×2 (experimental condition \times set size) repeated-measures design with the number of test examples counted correctly as the dependent measure. In all 32 trials the preliminary stage of the training was successful. All trials were therefore available for the statistical analysis. Since the Mauchly’s test of sphericity indicated the violation of the sphericity assumption for the condition factor ($\chi^2 = 28.918$, $p < 0.001$), lower-bound correction was applied in its case. The ANOVA indicated the main effect of the experimental condition to be statistically significant ($F = 542.167$, corrected $p < 0.001$, $\eta_p^2 = 0.946$). For the assumed confidence level $\alpha = 0.01$, the main effect of the set size was not statistically significant ($F = 32.505$, $p = 0.023$), but the interaction between the two within-subjects factors was ($F = 91.130$, $p < 0.001$, $\eta_p^2 = 0.401$). The planned contrast between the conditions indicated that the differences between both pairs of conditions (V^+G^+ versus V^+G^- , and V^+G^+ versus V^-G^+) were significant ($F = 341.238$, $p < 0.001$, $\eta_p^2 = 0.917$, and $F = 495.954$, $p < 0.001$, $\eta_p^2 = 0.941$, respectively). In the absence of the main effect of the set size, the corresponding contrast was of course also not

significant ($F = 5.747$, $p = 0.023$). The overall counting accuracy of the model (without the small/large numbers distinction) across the experimental conditions is shown in figure 25a. The plot of the estimated marginal means for the condition \times set size interaction is shown in figure 26.

The presence of the interaction between the experimental condition and the set size factors was followed by a post-hoc analysis of the significant differences. All possible 15 pairwise comparisons were examined using the paired-samples t -test, with the assumed confidence level $\alpha = 0.01$ adjusted for multiple comparisons using the Holm-Bonferroni method. This indicated that all differences were statistically significant ($p < 1 \cdot 10^{-5}$) except for two: small versus large sets in the V^+G^+ condition ($t = -0.242$, $p = 0.811$, adjusted $\alpha = 0.01$), and small versus large sets in the V^+G^- condition ($t = -2.665$, $p = 0.012$, adjusted $\alpha = 0.005$, see figure 26). The discovered interaction turned out therefore to be of the ordinal type.

Similarly to the analysis of the counting errors made by children conducted by Alibali and DiRusso (1999), the results of which are reproduced in figure 27, table 2 reports the percentage of the trials with particular types of errors as well as the percentage of models (out of 32) which made particular kinds of errors at least once, across the experimental conditions.

For each trial the actual counting duration was regressed against the counted set size for all examples in the test dataset. The obtained values of the regression slopes across the conditions are summarised in table 3.

6.3.4 Discussion

The statistical analyses indicated that the model’s counting accuracy differed between the experimental conditions. More sets in the test data were counted correctly by the models trained in the V^+G^+ condition (i.e. with the gestures) than by those trained in the V^+G^- condition (without the gestures). As indicated in figure 25a, the increase in the overall counting accuracy between these two conditions was quite

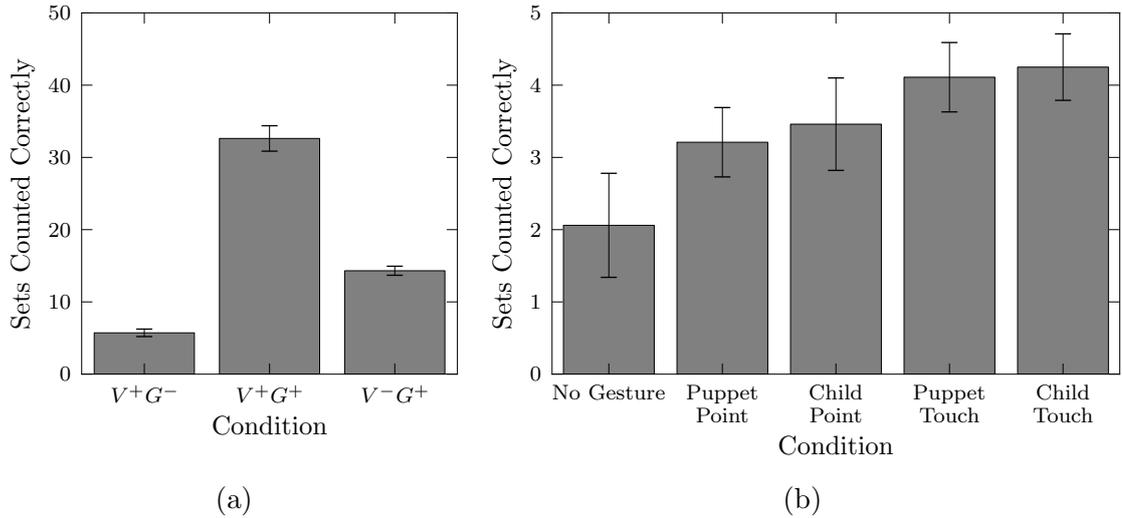


Figure 25: Overall counting accuracy of the model in simulation 3 across the experimental conditions (a), compared with human data (b). Error bars show 95% confidence intervals. All pairwise comparisons in (a) are statistically significant. In (b), ‘No Gesture’ condition is significantly worse than the ‘Puppet’ conditions combined and than the ‘Child’ conditions combined; ‘Point’ condition is significantly worse than ‘Touch’ condition for both ‘Child’ and ‘Puppet’; combined ‘Child’ was not significantly different from combined ‘Puppet’. Figure (b) reproduced from *Cognitive Development*, 14(1), Alibali, M. W. & DiRusso, A. A., *The function of gesture in learning to count: More than keeping track*, p. 46, Copyright 1999, with permission from Elsevier. The figure was modified to show 95% confidence intervals instead of standard errors, i.e. the error bars are twice as long as in the original.

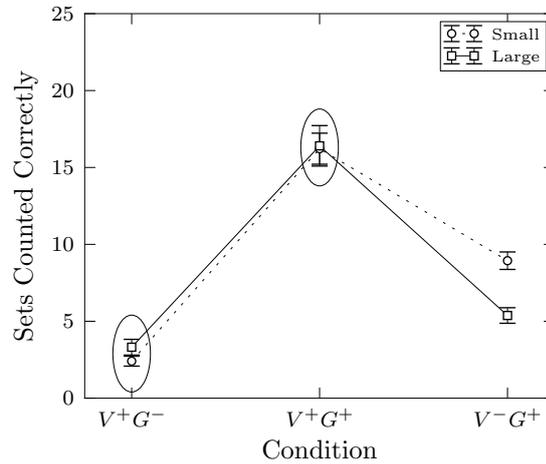


Figure 26: Profile plot for the simulation 3 ANOVA. The plot shows the number of small and large sets from the test data set (out of 25) counted correctly by the model after the second stage of the training in each condition, and illustrates the interaction between the experimental condition and the set size. The error bars indicate 95% confidence intervals. All possible pairwise comparisons except for the two encircled ones were significant at $\alpha = 0.01$.

	Percent of Trials with Error Made			Percent of Models which Made Error		
	V^+G^-	V^+G^+	V^-G^+	V^+G^-	V^+G^+	V^-G^+
	($N = 1600$)	($N = 1600$)	($N = 1600$)	($N = 32$)	($N = 32$)	($N = 32$)
Partitioning Errors						
Skip	0.0%	0.0%	0.0%	0%	0%	0%
Double count	NA	NA	NA	NA	NA	NA
Coordination Errors						
Continue	43.75%	17.86%	27.5%	100%	100%	100%
Stop short	44.81%	16.88%	43.88%	100%	100%	100%
Other Errors						
String Error	0.0%	0.0%	0.0%	0%	0%	0%
Distracted	NA	NA	NA	NA	NA	NA

NA = not applicable.

Table 2: Summary of the counting errors made by the model.

	Percent of Trials with Error Made		Percent of Children who Made Error	
	Child Conditions ($N = 200$)	Puppet Conditions ($N = 200$)	Child Conditions ($N = 20$)	Puppet Conditions ($N = 20$)
Partitioning Errors				
Skip	14.0%	NA	60%	NA
Double count	7.5%	NA	45%	NA
Coordination Errors				
Continue	0.5%	19.5%	5%	80%
Stop short	1.5%	2.5%	15%	25%
Other Errors				
String Error	5.5%	9.0%	30%	30%
Distracted	0.5%	0.0%	5%	0%

NA = not applicable.

Figure 27: Errors made in Child Conditions (Child Touch and Child Point) and Puppet Conditions (Puppet Touch and Puppet Point). Reprinted from *Cognitive Development*, 14(1), Alibali, M. W. & DiRusso, A. A., *The function of gesture in learning to count: More than keeping track*, p. 48, Copyright 1999, with permission from Elsevier.

Condition	Minimum	Median	Maximum
V^+G^-	-0.1	0	0.23
V^+G^+	0.84	0.89	0.95
V^-G^+	0.6	0.73	0.83

Table 3: Ranges of values of the regression slopes across the 32 trials obtained in the output length analysis.

dramatic, meaning that the proprioceptive signal provided the model with a significant amount of additional information which made it easier for the neural network to correlate its inputs with the desired output. This important result constitutes the evidence that the counting gestures provide useful information in the context of the counting task, which can be extracted even using simple learning mechanisms. Consequently, it provides a positive answer to the research question 2, stated at the beginning of this thesis. It is important to stress that this result was obtained despite the proprioceptive information being provided in a primitive form derived directly from the values of the arm joint angles. No explicit processing of this signal took place, which would facilitate extracting more abstract knowledge from the gestures (which in fact is likely to take place in the brain, at later processing stages), e.g. one that would highlight the spatial correspondence between the arm posture and the location of the object being pointed to. Nevertheless, the proposed model still managed to make good use of the additional proprioceptive information, what suggests that the pointing gestures are a useful embodied cue when learning to count.

The above finding is, in the qualitative sense, in perfect agreement with that reported by Alibali and DiRusso (1999) where the children's performance in the no-gestures condition was significantly inferior to all conditions which allowed the gestures (see figure 25b). As the comparison between the figures 25a and 25b reveals, the increase in the counting accuracy achieved in the present simulation is more prominent in terms of magnitude than that observed in children around 5 years of age, but, considering the relative simplicity of the proposed model, it would be unreasonable to expect it to fit the behavioural data exactly.

The very low counting accuracy achieved in the V^+G^- condition, contrasted with much better performance after the gesture signal has been added, raises a question whether the visual input is exploited by the model at all, in other words, if the improvement in the counting accuracy in the V^+G^+ condition could be attributed solely to the newly available proprioceptive signal. As mentioned earlier, this issue was addressed by the third experimental condition, V^-G^+ . As it is evident from fig-

ure 25a, and from the conducted contrast, the counting accuracy of the model in the V^-G^+ condition was significantly lower than that achieved in the V^+G^+ condition. Therefore, the dramatic increase in the counting competence in the V^+G^+ condition compared to V^+G^- must be a result of some kind of *fusion* of the information from both visual and proprioceptive inputs, as it is not explained by the gesture signal alone.

The proposed model failed to reproduce the effect of the set size reported by Alibali and DiRusso (1999). Significantly better counting accuracy on small sets than on large ones was obtained only in the artificial V^-G^+ condition, as indicated by the discovered interaction between the condition and set size factors (figure 26). This suggests that the extent to which the results of the model can be compared with human data is limited.

The additional insight into the differences between the behaviour of the proposed model and that of humans comes from the analysis of the frequencies of the counting errors (table 2 and figure 27). According to the behavioural study of Alibali and DiRusso (1999), the most common errors that children make when they count on their own are the partitioning errors (i.e. skip and double count). The error patterns obtained using the model are more similar to those of children in the puppet conditions (which, as discussed in section 6.3.2, are more appropriate for such a comparison) as there children make much more coordination errors, and continue errors become the most frequently committed ones. As it turns out, the counting errors committed by the proposed model were *strictly limited* to the coordination errors (continue and stop short). Every trained neural network in all 32 independent repetitions made both these errors, at least once, in all three experimental conditions. In contrast, none of the networks committed a single string or skip error in all trials and across all experimental conditions. Clearly, this indicates the existence of some inherent limitation of the proposed model in terms of the kinds of the counting errors it is likely to commit.

Theoretically, the possibility of skip and string errors is ruled out neither by the

architecture of the proposed model, nor by the assumed coding schemes of inputs and outputs. However, the following factors may have played a role in obtaining the error patterns reported in table 2. First, the Elman architecture, on which the model is based, treats time in a discrete fashion. As a consequence, the changes in the proprioceptive signal fed to the network are inherently synchronised with the production of the number words by the model. Connected with the assumption that the gesture is always correct (see section 6.3.2), this rules out the possibility of double count errors, as mentioned earlier. While skip errors are not ruled out (technically, the model could output ‘silence’ instead of a number word in the midst of counting, resulting in a skip error), it seems that the adopted discrete-time framework and the applied training regime makes it highly unlikely for the synchronisation errors to occur. This means that if the effects of the problems with synchronisation between action and speech in counting are to be simulated, a continuous-time model would most likely be more appropriate.

The second factor which may affect the types of the counting errors made by the proposed model is the verbal output representation. As reported in chapter 5, the model encodes the number words using one-hot coding. As a consequence, the representations of the number words are orthogonal, and therefore equally distant from each other in the representation space. This means that every word is equally likely to be confused with every other word, which could be the reason behind the fact that in the tests the networks never confused any two number words. One could hypothesise that if there was enough overlap between the representations of certain number words, they would be easier to confuse with one another and the string errors would be more likely to occur. It is not clear however on what level such an overlap should exist. One possible source of the representational overlap could be the phonological similarity between the words. On the other hand, the similarity could lie at the semantic level as well.

The regression slope values obtained in the V^+G^+ condition correspond well to correct counting (see table 3). This is not surprising, considering the good counting

accuracy achieved in that condition. In the V^+G^- condition however, the slope values are much lower and concentrated around 0, suggesting that the deteriorated behaviour described in section 6.3.2 may have appeared. A closer look at the actual output of the models in the V^+G^- condition across the testing dataset revealed however that, although a completely deteriorated behaviour actually never appeared (only in 1 trial the model counted always either to 5 or 6; in all other cases the model output was more complex), overall, the length of the model’s counting in this condition did not really depend on the size of the counted collection. This suggests that the task that was posed to the neural network was quite hard to solve using the applied training technique, based only on the visual input. This may be the consequence of the chosen visual input encoding, which intentionally does not contain obvious cues about the cardinality of the set being represented (see chapter 5). Interestingly, the fact that the counting accuracy obtained in the V^+G^+ condition was significantly better than both V^+G^- and V^-G^+ conditions, suggests that the proprioceptive signal enabled the model to make more sense of its visual input.

6.4 Simulation 4 — Significance of the Spatial Aspect of the Counting Gestures

6.4.1 Aims of the Experiment

Simulation 3 confirmed that providing the proposed model of learning to count with the proprioceptive stimuli in addition to the visual information enables it to achieve higher counting accuracy. The experiment did not however answer which characteristics of the gesture signal are responsible for the the observed improvement.

Some of the hypotheses about the nature of the contribution of the counting gestures to learning to count highlight the fact that such gestures correspond to how the counted items are distributed in space (see section 2.2.2). They link therefore

the spatial aspect of the enumerated set with the temporal structure of the count list in one motor activity. In the proposed model of learning to count, this is reflected in the following way. The activation values of the proprioceptive input units, as they are presented to the neural network over time, always unambiguously correspond to the locations of the objects in the visual input (cf. figure 13). Because it is known which arm postures correspond to which locations, it is actually possible, based on the proprioceptive information, to reconstruct the arrangement of the items in the visual input. At the same time, the gesture signal unfolds over time in a way that corresponds to the correct recitation of the number words in counting. Synchronisation of the production of the number words with the gesture signal leads therefore to correct enumeration. The counting gestures in the model may thus act as a link between the spatial characteristics of the counted set and the temporal characteristics of the sequential enumeration.

The present simulation aims to investigate if the counting accuracy of the proposed model is affected when the spatio-temporal link described above is broken. The results are expected to provide the basis for the answer to the third research question considered in this thesis. In order to understand better the importance of the investigated aspect of the counting gestures, in this experiment the proprioceptive representation is modified in such a way that it still carries the temporal information, but the spatial aspect, typical for the ‘natural’ counting gestures, is not present any longer.

6.4.2 Procedure

The experimental design in this simulation was analogous to that used in simulation 3. Following the preliminary training stage, which equips the neural network with the ability to recite the count list, the second stage of the training was performed in four experimental conditions. First two of the conditions replicated the V^+G^+ and V^-G^+ conditions from simulation 3. The other two conditions, which from now on will be referred to as $V^+G_R^+$ and $V^-G_R^+$, used the same neural network

configurations as the V^+G^+ and V^-G^+ conditions, respectively (see figure 28), but, instead of accepting the spatio-temporal proprioceptive information constructed as described in section 5.3.2.1, they were fed with the signal corresponding to the gesture representation introduced in section 5.3.2.2.

These temporal-only (or *rhythmic*, hence the G_R) gestures represent an activity in which the robot moves its arm back and forth for the number of times equal to the number of items in the enumerated set. In a proprioceptive signal constructed this way, the temporal aspect of the gestures matches the spatial aspect of the counted set (the number of performed movements is equal to the number of objects in the set), however the correspondence between the locations of the objects and where the arm is pointing throughout the gesture is not present anymore. In a study with children, this would correspond to allowing them to gesture while counting e.g. by tapping on the table, but not by pointing to the items being counted. In the puppet conditions, the puppet would jump (or perform another distinct rhythmic activity) for the number of times equal to the number of objects in the set.

Between the four experimental conditions in this simulation there are thus two orthogonal factors. One of them is the type of the gestures (spatio-temporal or rhythmic), and the other is the presence of the visual input in the model configuration. The differences in the counting accuracy as the result of a change in these two factors were the planned contrasts.

The training and evaluation of the neural networks proceeded exactly as in simulation 3. The same training parameters were used, N_H was also fixed to 20, and 32 independent repetitions of the experiment were performed. The distinction between the small and large sets was however no longer made. A parameter of the rhythmic gestures, which may have an effect on the ability of the neural network to exploit the information it carries, is the *amplitude* of the movement. Since it is difficult to argue what amplitude of the rhythmic gesture would convey the same ‘amount of rhythmic information’ as the spatio-temporal gestures, a rhythmic gesture with maximum possible amplitude was used ($l = 1$ and $r = 20$).

6.4.3 Results

As mentioned above, the four experimental conditions in this simulation form two 2-level orthogonal factors, what yields a 2×2 (type of gestures \times presence of vision) repeated-measures design. The dependent measure was, as in previous simulations, the number of sets from the test data set counted correctly. In all 32 trials the preliminary stage of the training completed successfully, allowing all trials to be included in the analysis. Both planned contrasts (spatio-temporal versus rhythmic gestures and vision present versus absent) turned out to be statistically significant ($F = 231.077$, $p < 0.001$, $\eta_p^2 = 0.882$ and $F = 122.358$, $p < 0.001$, $\eta_p^2 = 0.798$, respectively), however there was also a significant interaction between these factors ($F = 178.140$, $p < 0.001$, $\eta_p^2 = 0.852$). This interaction is illustrated in the profile plot shown in figure 29.

The discovered interaction prompted a post-hoc analysis in order to find out where the statistically significant differences in the counting accuracy were. As in previous cases, this was done using the paired-samples t -test, adjusted for multiple comparisons using the Holm-Bonferroni method, and assuming the family-wise confidence level $\alpha = 0.01$. Out of 6 possible pairwise comparisons, 5 were statistically significant ($p < 1 \cdot 10^{-7}$). The only comparison in which the difference in the counting accuracy was not statistically significant was $V^+G_R^+$ against $V^-G_R^+$ ($t = -0.711$, $p = 0.482$, adjusted $\alpha = 0.01$, see figure 29). The interaction between the within-subjects factors was therefore of the ordinal type.

6.4.4 Discussion

The first important finding in the present simulation is that the performance of the model in terms of counting accuracy in the conditions with the rhythmic gestures ($V^+G_R^+$ and $V^-G_R^+$) was higher than that achieved in the condition with the ‘natural’ counting gestures (V^+G^+). At first glance, this suggests that in simulation 3 it was the temporal (rhythmic) aspect of the gestures that contributed to the increase in

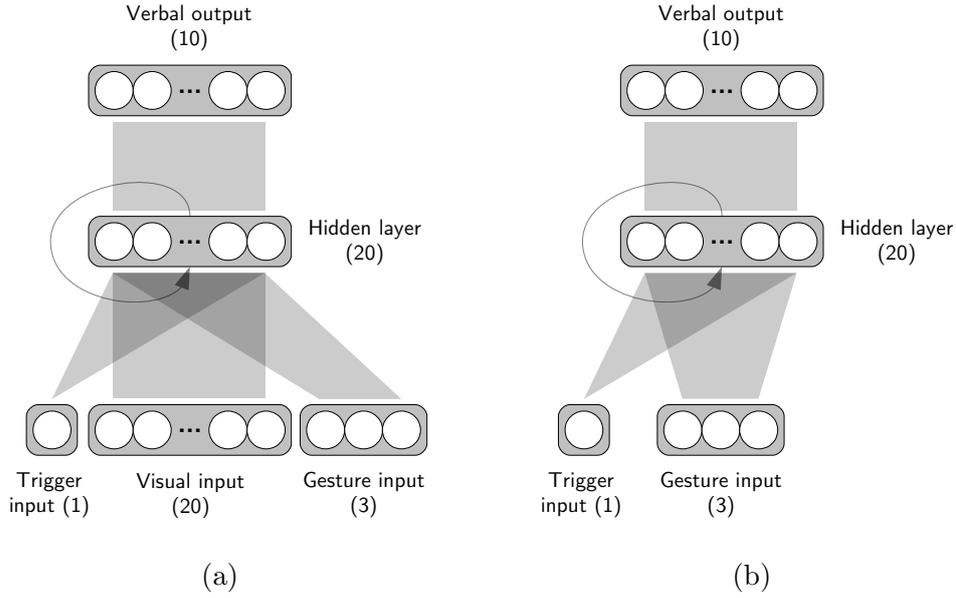


Figure 28: Configurations of the model used in simulation 4. Configuration (a) was used in the V^+G^+ and $V^+G_R^+$ conditions, configuration (b) in the V^-G^+ and $V^-G_R^+$ conditions. In all cases $N_{VO} = 10$, $N_{VI} = 20$, $N_G = 3$, and $N_H = 20$.

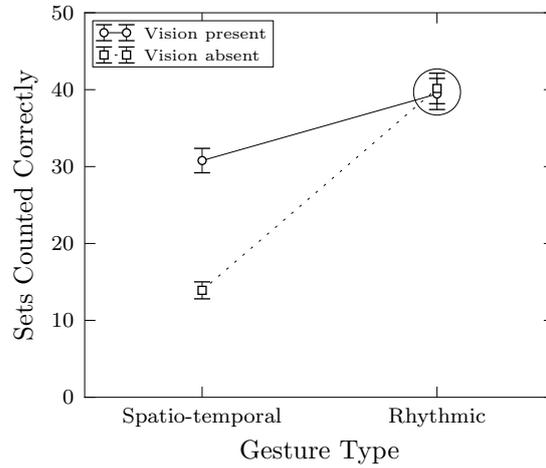


Figure 29: Profile plot for the simulation 4 ANOVA. The plot shows the number of collections from the test data set (out of 50) counted correctly by the model after the second stage of the training, illustrating the interaction between the gesture type and the presence of vision factors. The error bars show 95% confidence intervals. All pairwise comparisons, except for the one encircled, were significant at $\alpha = 0.01$.

the counting accuracy of the model in the V^+G^+ condition (compared to the V^+G^- condition), because when this aspect of the gestures was made even more prominent in the $V^+G_R^+$ (and $V^-G_R^+$) condition of the present simulation, this led to even higher counting accuracy. A deeper reflection on the results from all four conditions included in the present simulation reveals however that the real picture is slightly more complicated.

An important difference between how the spatio-temporal and rhythmic gestures were exploited by the proposed neural network is revealed by the comparison between the corresponding pairs of conditions which use the same type of gestures — V^+G^+ versus V^-G^+ and $V^+G_R^+$ versus $V^-G_R^+$. The discovered interaction between the presence of vision and the type of gestures means that, in contrast to the conditions which used the spatio-temporal gestures, the difference in the counting accuracy between the conditions which used the rhythmic gestures was too small to show up as statistically significant (cf. figure 29). Note, that this does not mean that the counting accuracy in these two conditions was identical — the difference was however significantly smaller than for the V^+G^+ versus V^-G^+ pair. In other words, when provided with the proprioceptive information in the form of rhythmic gestures, the model achieved very similar counting accuracy with and without the visual input. This however indicates, that the high counting accuracy observed in the $V^+G_R^+$ condition can be explained by the contribution of the proprioceptive information alone — therefore, when the rhythmic gestures were used, the contents of the visual input were, apparently, ignored. The rhythmic proprioceptive signal enabled the model to achieve high counting accuracy overall, but it did not enable it to extract more information from the visual input. This is opposite to what has been found for the spatio-temporal gestures in the present and previous simulation, where providing the model with both proprioceptive and visual information (V^+G^+) significantly improved the counting accuracy over the gesture-only condition (V^-G^+), what suggests reliance on both sources of information in the former case. The interaction between the presence of vision and the type of gestures provides thus the evidence

that there is a crucial qualitative difference in how the ‘natural’ gesture signal, which combines the spatial and temporal aspects of counting in one motor activity, and the ‘rhythmic’ gesture signal, which does not link these two aspects, affect the process of learning to count in the proposed model. Even though the latter type of gestures resulted in higher counting accuracy overall, it did not help the model to improve its counting ability in the sense that the contents of the visual input, which after all contains the items to be counted, were in fact ignored. This computational modelling evidence suggests that the answer to the research question 3 stated in section 1.2 is positive.

The results obtained with the rhythmic gestures are more easily interpreted when one considers an analogous situation in a behavioural study. If one would like to replicate the above results in a ‘passive rhythmic gesture condition’, the child could be asked to count items on the table while a puppet would perform a rhythmic activity, for example jumping. Since the correspondence between the items on the table and the movements of the puppet would be indirect (the number of jumps would match the number of items on the table; note however that recognising this would require the child to already have counted the items) it is likely that the child would revert to counting the puppet moves rather than the objects on the table. In fact, a retrospective analysis of the experimental data gathered by Alibali and DiRusso (1999) revealed that the children in this study had such a tendency even for the spatio-temporal gestures made by the puppet. When commenting on the level to which the children were engaged in the experiment, Alibali and DiRusso report:

Indeed, at least three lines of evidence suggest that children were attentive and engaged in the puppet conditions. First, when the puppet counted incorrectly (in the puppet-incorrect condition), children almost always counted incorrectly as well. In fact, *their counts most often coincided with the number of indicating acts produced by the puppet, and not with the actual number of chips.* (Alibali & DiRusso, 1999, p. 52, italics ours)

Therefore, the fact that in the $V^+G_R^+$ condition the neural network was counting the number of the robot arm swings, rather than the items in the visual input, is actually

plausible from the behavioural point of view. Unfortunately, Alibali and DiRusso do not provide quantitative data, that could be compared with the modelling results.

6.5 Summary

In chapter 5 of this thesis I described a computational model of learning to count designed according to the developmental cognitive robotics paradigm, the main purpose of which is the investigation of the contribution of the counting gestures to the acquisition of the counting skill. Although based on a long-existing artificial neural network architecture, the model contributes to the state-of-the-art in the modelling of mathematical cognition being one of the very few models which incorporate all crucial components of learning to count (vision, speech and gestures), but also, and more importantly, because it is the first one to employ a realistic (from the point of view of embodied cognition) representation of the proprioceptive information, constructed with the use of an artificial humanoid body of the iCub robot.

In the present chapter the model has been subjected to a series of simulation experiments, which, first of all, confirmed the validity of the proposed solution in terms of the fundamental criterion used in machine learning, namely the ability to generalise — that is to transfer the desired behaviour from the input patterns on which the neural network has been trained to the novel ones which have never been encountered by the model before. In addition, the results of the simulations were helpful in finding good values of certain parameters of the model as well as of the associated training regime. More importantly however, the simulation experiments provided the answers to the first three of the research questions considered in this thesis (see section 1.2). In simulation 2 it has been shown that explicitly dividing the training of the model into two stages — the acquisition of the count list followed by learning to count — which is inspired by the psychological knowledge about the development of the counting skill in children, makes it possible to achieve higher counting accuracy within the same amount of training effort than requiring

the model to master both skills at once. This suggests that mastering the count list prior to learning to count within the respective range of collection sizes may speed up the subsequent process of learning to count. Simulation 3 provided the first evidence for the usefulness of the proprioceptive information conveyed by the counting gestures in the context of learning to count based on computational simulations and not on behavioural data. The way the gestures contributed to the counting accuracy was not trivial, as the improvement could not be explained by the presence of the proprioceptive signal alone. Since the employed representation of the counting gestures was based on the values of arm joint angles changing over time, this proves that even in such primitive form, counting gestures are a useful embodied clue which can easily be exploited to improve the counting accuracy. Finally, the results of the simulation 4 demonstrated the importance of the spatial correspondence between the items being enumerated and the indicating act performed during counting by showing that natural counting gestures, characterised by such a correspondence, help the model to make better use of the visual information, while this is not the case for the rhythmic gestures, which carry only the temporal information and in which such a correspondence does not exist.

Answering the final, fourth research question addressed in this thesis requires adopting a slightly different point of view at the process of the development of human numerical skills, as a more extensive time scale needs to be considered. The subsequent chapters present a developmental model and simulation experiments which focus on the investigation of the ontogeny of the spatial-numerical associations.

Part III

Neuro-Robotic Model of the Acquisition of Spatial-Numerical Associations

Chapter 7

Model Overview

In this part of my thesis I turn my attention to another set of embodied phenomena connected with learning to count, namely to spatial-numerical associations (including the SNARC effect, see section 2.2.3). From the developmental perspective, this second set of modelling experiments focuses on a broader extent of the time line of the acquisition of the numerical knowledge than the first one. While in the previous two chapters I have looked at the developmental phenomena that take place in a relatively short period of time, during which the children acquire the ability to count correctly, here a wider perspective is adopted, as the longer-term effects that the way we learn to count has on our behaviour in numerical tasks are considered.

In this chapter I propose a neuro-robotic model of the acquisition of spatial-numerical associations, with the intention to investigate the role of the cultural and environmental factors in the establishment of a connection between these two concepts. More specifically, I aim at demonstrating that the systematic spatial biases that children are exposed to when learning to count can lead to the emergence of certain behavioural effects found in human performance in simple numerical tasks later in life (cf. research question 4). Similarly to the structure of chapter 5, I first present the assumptions behind the undertaken modelling effort, and then proceed to the description of the model itself and of the development process associated with it. The simulations conducted using the model are the topic of the subsequent chapter.

7.1 Model Design Assumptions

As reviewed in chapters 2 and 3, spatial-numerical associations not only have drawn a great amount of attention on the part of the experimental psychologists, but have also been the subject of computational modelling efforts in the past. The main focus of the previous computational models of the interactions between numbers and space was to explain the mechanisms behind the behavioural effects these interactions are evident in. In contrast, here more emphasis is put on the reasons why and how spatial-numerical associations may be acquired.

Considering the existence of well-established models of the spatial-numerical associations (Gevers, Verguts et al., 2006; Chen & Verguts, 2010), the present study is solidly rooted in these efforts, to the point that certain parts of the model presented here are a reproduction of the work of the quoted authors. My unique contribution is connected with the shift of the centre of attention to the process of the development of the model and with the enrichment of the representation of embodied spatial cognition. The proposed training regime throughout which the model is obtained is important from the theoretical point of view. Significant attention has been given to the degree to which it resembles the ontogeny of numerical knowledge in children and to its link with sensorimotor development.

In addition, as was the case with the model of learning to count presented in the previous part of this thesis, an important novel aspect of the adopted modelling methodology is connected with the inclusion of an artificial robotic body in the process. Also here such an approach is appropriate, since the phenomena being considered have a definite embodied character.

The primary hypothesis behind the present experiment is that it is the tendency of children to learn to count from left to right that may be responsible for the emergence of the association of small numbers with the left side of space and large numbers with the right side of space. This assumption is reflected in the way the development process of the model is designed. Subsequently, in simulations of the numerical tasks, it is investigated if, after the model is developed, it exhibits the

patterns of behaviour which are widely accepted as the manifestations of spatial-numerical associations.

The following three tasks are considered in the behavioural simulations of the model: number magnitude comparison, number parity judgement and visual target detection task. As reviewed in chapter 2, the former two tasks are extensively used in the experimental study of human numerical skills, and different well-established effects can be assessed using these. More specifically, number comparison is a classical paradigm, in which the number size and the numerical distance effects are found (see section 2.1.1). The same task is also commonly used in the investigation of the SNARC effect (see section 2.2.3), as is the parity judgement task. The latter is especially important in this context, since in its case the magnitude of the processed number is irrelevant. Finally, the visual target detection task, in which the Posner-SNARC effect is found (see section 2.2.3), demonstrates the effects of spatial-numerical associations despite the fact that in this task the subjects are not expected to perform any kind of numerical processing explicitly. Taken all together, a successful emergence of the appropriate patterns of behaviour in those tasks should suffice to demonstrate that the proposed model has indeed acquired the spatial-numerical associations.

7.2 Model Architecture

The architecture of the proposed neuro-robotic model of the acquisition of spatial-numerical associations extends the work of Chen and Verguts (2010, see chapter 3, section 3.4). Additionally, some of the employed solutions were inspired by the principles used by Caligiore, Borghi, Parisi and Baldassarre (2010) in formulating their model of compatibility effects, which focused on motor affordances and goals. The architecture of the model proposed in this chapter is shown in figure 30.

On the general level, the proposed neural network adopts an approach of two pathways, which has been successfully used in the context of compatibility effects

before (Gevers, Verguts et al., 2006; Caligiore et al., 2010), and which takes inspiration from a high-level organisational principle hypothesised to exist in the areas of the brain responsible for visual processing (Ungerleider & Mishkin, 1982; Milner & Goodale, 2008). According to this highly influential theory, the processing of the visual information in the brain follows two mostly independent streams, called *ventral* and *dorsal*. Without delving into the neuroanatomical details, the ventral stream processes the information about the identity of the objects (and therefore is sometimes referred to as the *what* stream), whereas the dorsal stream is concerned with the objects' locations (and thus called the *where* stream). As indicated in figure 30, the proposed architecture is built according to an analogous organisational principle. However, it is important to remark that, just as it was the case for the model of counting described in chapter 5, it is not my aim herein to formulate a neurophysiological model of the considered phenomena, in which particular components would correspond to well-defined areas of the human brain. From now on therefore, the terms 'Where pathway' and 'What pathway' used in the present work should be understood figuratively, as referring to the way the processing of the information in the model is organised; they should not be interpreted as any form of a claim that the model refers to some specific aspects of the brain physiology. Below I discuss both pathways of the proposed model and provide the details about the representational elements they consist of.

7.2.1 'What' Pathway — Symbolic Processing and Decision Making

The What pathway of the proposed neural network is designed to process symbolic information, which includes semantic encoding, decision making and the selection of the response. This part of the model is essentially a reproduction of the corresponding components of the model by Chen and Verguts (2010), which is being extended herein. Just as in the quoted work, the What pathway is a feed-forward network, consisting of four layers of units — input layer, semantic layer, decision layer and

response layer — in which the activations propagate from the input layer toward the response layer (see figure 30).

The *input layer* realises the input of number information to the model. It is assumed that numbers are conveyed using a symbolic format, such as the Arabic notation or number words. This layer employs the one-hot coding representation, in which every symbol corresponds to a configuration of activations in the layer with exactly one unit activated and all other units in the layer deactivated. Since in this scheme all possible input vectors are orthogonal, it is not important which particular unit in the layer represents which number. For convenience, it is therefore assumed that consecutive units in the layer correspond to consecutive numbers. The number of units in the input layer N_I is a parameter of the model.

The layer that follows the input layer, the *semantic layer*, represents the number identity or, in other words, the semantics of the symbol provided at the model's input. This layer implements the magnitude-based representation of numbers hypothesised to be employed in the brain in the form of a *mental number line*. Although quite a variety of encoding schemes for this kind of number representation have been proposed and evaluated (see section 3.2.2), in line with the approach of Chen and Verguts (2010), in the present work the coding with linear scaling and constant variability is used. Linear scaling means that the location of the most strongly activated unit in the semantic layer is proportional to the magnitude of the number being represented (an alternative is to use coding that is compressed, e.g. logarithmically). Constant variability means that the representations of the nearby numbers overlap with each other to some extent (i.e. in the pattern of activations representing the number 5, units which represent the numbers 4 and 6 are activated to a certain degree too) and moreover, this overlap is the same irrespective of the absolute number magnitude (an alternative are representations with scalar, that is increasing variability). In the adopted encoding, if the number k is being represented, the activation value of the unit i in the semantic layer is assumed to be $e^{-|k-i|}$. This encoding is implemented by the weights of connections between the input and semantic layers

(see figure 31). The semantic layer is assumed to consist of the same number of units as the input layer.

The activations in the semantic layer are used by the *decision layer* to execute simple numerical tasks commonly used in the behavioural study of spatial-numerical associations (see section 2.2.3). In the simulations conducted using the model in chapter 8, two numerical tasks are considered: number comparison and parity judgement. Both these tasks are realised using a decision layer with two units. In the number comparison task, the units correspond to the decisions ‘the first number is larger’ and ‘the second number is larger’. In the parity task, the interpretation of the units are ‘the number is odd’, and ‘the number is even’. The weights of the connections between the semantic layer and the decision layer are obtained in the course of supervised training, which is described later in section 7.3.

Finally, the *response layer* integrates information from the decision layer with those coming from the Where pathway, and is responsible for the final selection of the motor response. This layer consists of two units, representing two possible motor responses (e.g. left or right button push). The response given by the model is determined based on the activation values of the units in this layer as described in section 7.4.

7.2.2 ‘Where’ Pathway — Spatial Coding and Transformations

The Where pathway is the part of the proposed neural network that distinguishes it from the previous attempts to model spatial-numerical associations (Gevers, Verguts et al., 2006; Chen & Verguts, 2010). In contrast to these works, the advantage of the tools provided by the developmental neuro-robotics paradigm is taken here, which makes it possible to formulate embodied hypotheses about the way the spatial-numerical associations are acquired. The design of the Where pathway takes inspiration from the hypothesis that human spatial cognition employs numerous interacting spatial maps, which encode locations using different frames of reference

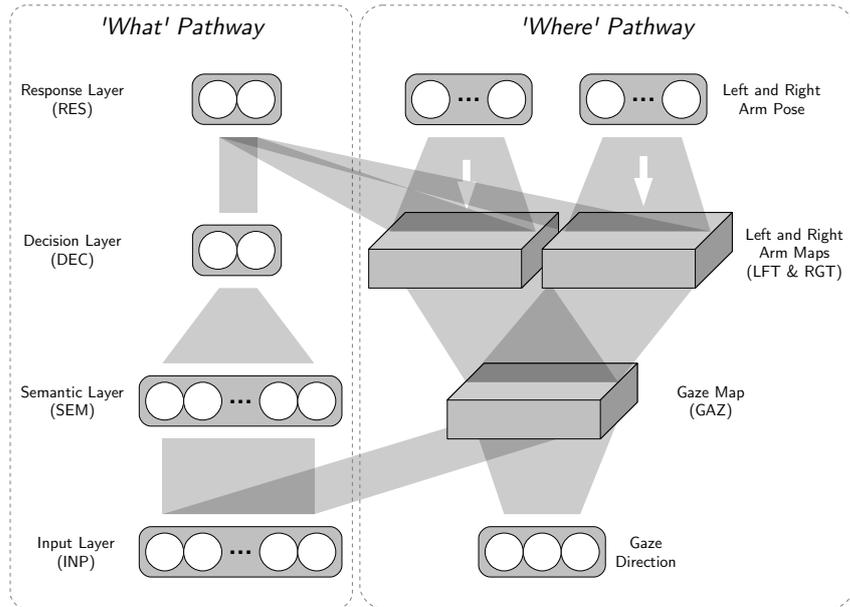


Figure 30: The neural network architecture of the model of the acquisition of spatial-numerical associations. Gray areas represent all-to-all connections. Rectangular blocks represent clusters of units forming self-organising maps. Activations in the network propagate from bottom to top, except for the connections with explicit indications.

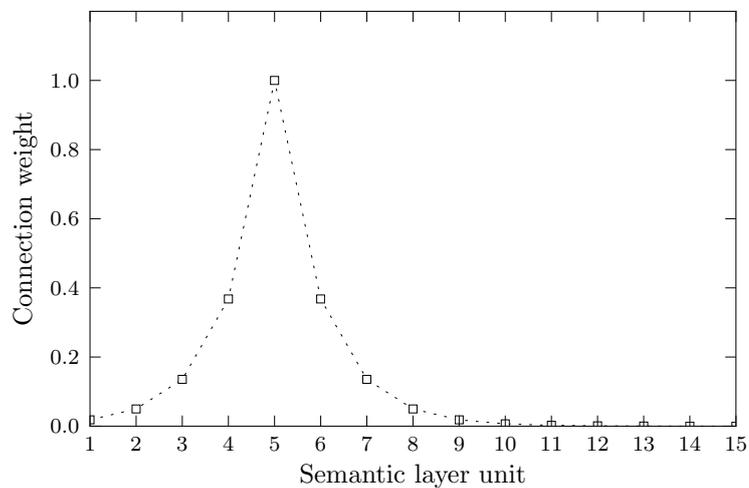


Figure 31: Weights of the connections between the input and the semantic layers, which implement a mental number line with linear scaling and constant variability, assuming $N_I = 15$. The abscissa corresponds to the index of the semantic layer unit, the ordinate shows the connection weight. The plot shows the weights of the connections from the input layer unit number 5 to all units in the semantic layer. Therefore, the plot illustrates the asymptotic distribution of the activations in the semantic layer when number 5 is being represented.

(Newcombe & Huttenlocher, 2000; Wang, Johnson, Sun & Zhang, 2005). In the proposed model, there are three neural maps, which represent the peripersonal working space of the iCub humanoid robot (cf. figure 30): one connected with the robot’s gaze direction, and two associated with the robot’s arms. These maps, implemented using Kohonen’s Self-Organising Maps (SOMs, see section 4.4.3), can be seen as performing both dimensionality reduction and sparse coding. The input to the gaze map is a 2-dimensional proprioceptive vector representing the iCub’s gaze direction (azimuth, and elevation angles). The input to each of the arm maps consists of a 7-dimensional proprioceptive vector representing the position of the relevant arm joints, namely shoulder pitch, roll, and yaw, elbow angle, and wrist pronosupination, pitch, and yaw. All three maps have 2-dimensional output topology. As a result, spatial locations around the robot expressed in the visual and arm-related frames of reference are represented in the neural network as 2-dimensional patterns of activations in the maps.

The gaze and arm maps are arranged in a stack, with all units of one map connected with all units of the other one with parallel, bi-directional links (see figure 30 and section 4.4.3). This set-up implements the transformation of coordinates between the frames of reference corresponding to the maps. The pattern of activations in the gaze map can be propagated to an arm map, what makes it possible to find the arm posture corresponding to reaching to the activated visual location — thus implementing a simple solution to the inverse kinematics problem. Propagating the activations from an arm map to the gaze map makes it possible to direct the gaze of the robot to where its arm end effector is located. This implements forward kinematics.

The hypothesis about the possible embodied sources of the SNARC effect is expressed in the design of the proposed model as the way in which the What pathway links with the Where pathway. As indicated in figure 30, in the proposed neural network, the symbolic input may be associated with the spatial locations in the visual frame of reference. This is intended to represent the fact that numbers may become

associated with space, for instance as the result of systematic spatial biases present during the development of number knowledge (see section 2.3.3). In addition, it is assumed that the activations in the spatial representations connected with arms may prime left- and right-sided motor responses. This is resembled in the model by the links between the arm maps in the Where pathway and the response layer in the What pathway.

7.2.3 Model Implementation

The simulations presented in chapter 8 use the model in three behavioural tasks, each having slightly different requirements. For instance, the parity judgement task requires one number to be given at input, while the number comparison task requires two operands. In turn, the simulation of the Posner-SNARC effect does not involve the decision layer at all. This issue was addressed in a similar way as in the previous modelling studies (Chen & Verguts, 2010; Grossberg & Repin, 2003). Depending on the task at hand, the actual structure of the model was adapted to the current needs by removing irrelevant components or by duplicating some others in order to implement short-term memory. The particulars are provided along with the description of each simulation in chapter 8.

Since the simulations focus on the measurement of the response times (RT), the neural network is implemented using the firing rate model (see section 4.4.2). Based on the connectivity of the neural network, the activity of each unit in the model is described by an ordinary differential equation, which expresses the change of the activity of the unit over time as a function of its inputs. The activation values of the units in the model can then be computed for a given set of inputs to the network using numerical integration. The latter was realised using the VODE integrator from the SciPy library for the Python programming language (Oliphant, 2007), with the integration time step equal to 0.01. The stop condition for the integration process was sufficient stabilisation of the activation values of all units (i.e. all derivatives must have fallen below $5 \cdot 10^{-5}$). The equations describing the

model in each considered configuration are given in appendix A.

The spatial maps in the Where pathway are implemented as 49-cell (7×7), 2-dimensional SOMs. In the gaze SOM topology the units are arranged in a square tiling, while the arm maps use the triangular tiling. The neighbourhood function used for training the SOMs is the Gaussian function, given by:

$$\Theta(i, j, t) = e^{-\frac{d_{out}(i, j)^2}{2\sigma^2(t)}} \quad (7.1)$$

where $d_{out}(i, j)$ is the distance between units i and j in the SOM output space, and $\sigma(t)$ is a parameter determining the spread of the neighbourhood. The activation function used to calculate the activation value of the SOM unit i given the input vector \mathbf{x} is the following exponential function:

$$y_i(\mathbf{x}) = e^{-\frac{\log(y_{min})\|\mathbf{x} - \mathbf{w}_i\|}{d_{norm}}} \quad (7.2)$$

where y_{min} and d_{norm} are scaling parameters. The behaviour of the function $y_i(\mathbf{x})$ is such that $y_i(\mathbf{x}) = 1$ for $\mathbf{x} = \mathbf{w}_i$ and $y_i(\mathbf{x}) = y_{min}$ for $\|\mathbf{x} - \mathbf{w}_i\| = d_{norm}$. The value of y_{min} was assumed to be 0.001, while d_{norm} was different for the gaze and arm maps (see table 4).

7.3 Developmental Learning

Consistent with the postulates of developmental cognitive robotics, the process of training of the proposed neural network is an important part of the present modelling effort. The learning of the neuro-robotic model of the acquisition of spatial-numerical associations is organised in four phases, intended to resemble the stages of the development of numerical knowledge of a human child (see chapter 2). Considering the latter from the point of view of the elements included in the proposed model leads to the developmental sequence shown in figure 32. The spatial representations for visual and motor affordances are built and correspondences between them are

established in the early months of human life. Somewhat later, between 2 and 4 years of age, children learn number words and their meaning, through the process of learning to count. Finally, usually in the late preschool or early school years, children are taught simple numerical tasks, such as number comparison or parity judgement. How these stages of the development are reflected in the proposed model is explained in details below.

7.3.1 Building Spatial Representations and Transformations

This phase of the network training focuses on building the gaze and arm maps and the connections between them (figure 32a). The spatial maps are built based on the data obtained by simulating, using the iCub humanoid robot, the process of *motor babbling* (Von Hofsten, 1982). Through motor babbling, children are believed to refine their internal representations of space. This process involves, for example, performing random movements with the arms while observing the hands at the same time, or reaching for toys in one's visual field. This makes it possible to perform later in life such tasks as visually-guided reaching.

This stage of the development was simulated using the iCub humanoid robot by performing movements with the robot's arms and gaze within what was assumed to be the robot's operational space. This was chosen to be a section of a sphere with the centre placed between the robot's shoulder joints, the radius of 35 centimetres, and the span of $\pm 30^\circ$ in elevation and $\pm 45^\circ$ in azimuth in front of the robot (see figure 33). This sphere section was subdivided uniformly into a 21×7 grid, yielding 147 target locations for directing the gaze and moving the arms of the robot using the Cartesian controllers. In order to gather data about the arm and head angles corresponding to the same locations in the robot's operational space, simulated motor babbling proceeded in trials, in which the gaze of the robot, and both its arms were commanded to fixate on and reach to the same target point. The resulting gaze and arm postures were then read from the robot and stored. Between each trial, the head and arms of the robot were moved to a rest position, in order to eliminate

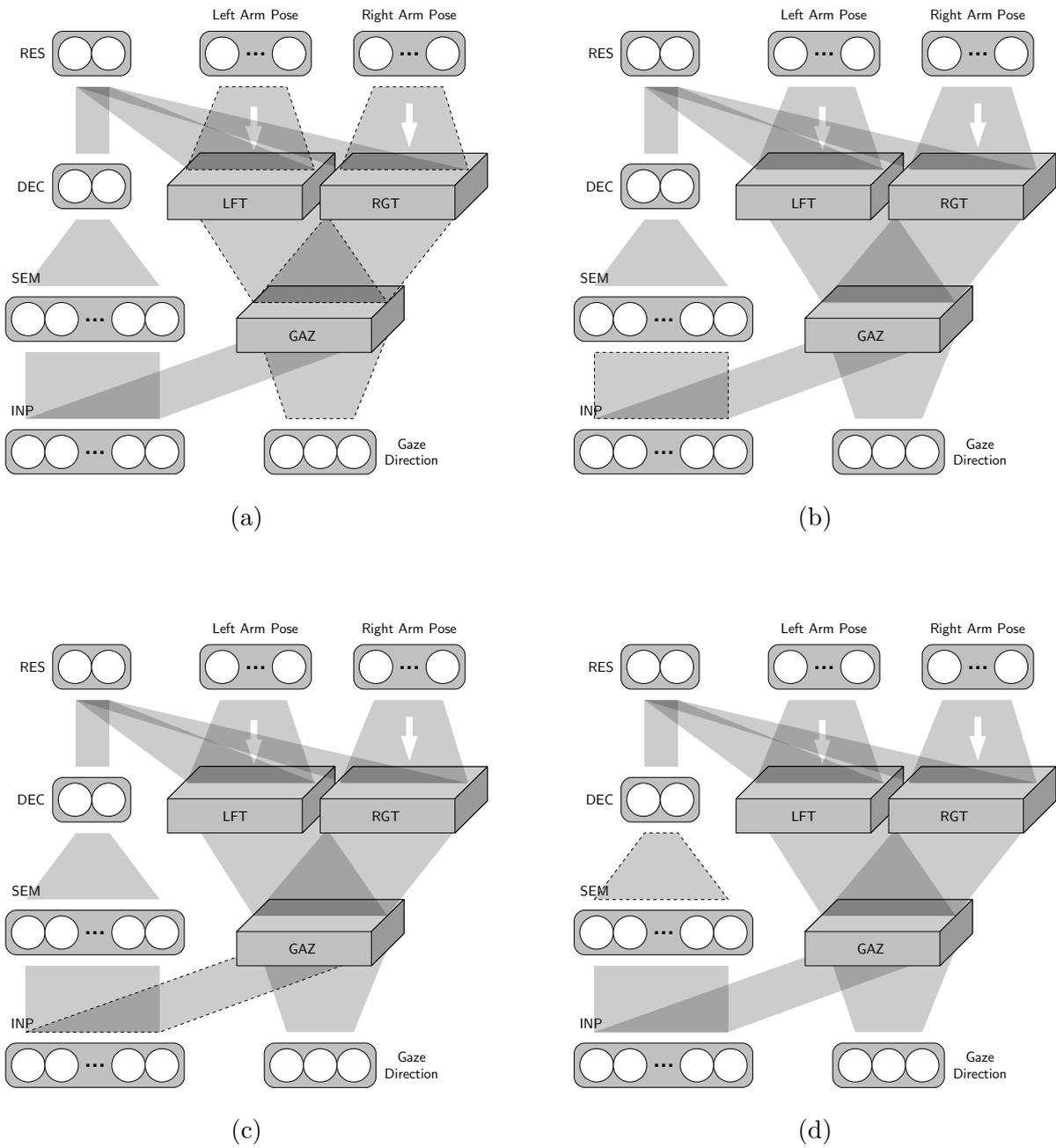


Figure 32: Development of the model of spatial-numerical associations. Connection weights being modified at each stage are marked with a dashed outline. (a) building spatial representations and transformations, (b) learning the meaning of number words, (c) learning to count, and (d) learning of numerical tasks.

any potential influence of the sequence in which the target points have been used on the resulting head and arm postures.

The joint angles data collected during motor babbling were subsequently used to construct the three SOMs, using the classical unsupervised training algorithm (see section 4.4.3). The gaze SOM was trained using all collected data. The arm SOMs were trained using the data coming only from these trials, in which the considered arm reached no further than 7.5cm from the target location. This reflects the natural morphological asymmetry between the reachable space for the left and right arm, which results in that some parts of the robot’s operational space are reachable by the right arm, but not by the left arm (and vice versa; see figure 34). The parameters of the SOM training algorithm were adjusted through trial-and-error, based on the observation of the learning process and the analysis of how well resulting networks span the target spaces. The final values of the parameters that were used to construct the SOMs are reported in table 4.

The transformations between the visual and arm-related frames of reference are implemented as the connections between the spatial maps (cf. figure 32a). The weights of these connections were obtained using to the classical Hebbian learning rule (Hebb, 1949). The vectors with gaze and arm angles corresponding to the same target point, obtained during the motor babbling, were fed to the appropriate SOMs, and the activation values of the units in these were calculated. The values of the connection weights were computed as a sum of the products of the activation patterns across the entire data set, and subsequently normalised in such a way, that the total activation propagated through the weights when exactly one unit of the pre-synaptic (i.e. the gaze) map is fully activated does not exceed 1.

7.3.2 Learning the Semantics of Number Symbols

This stage of the development corresponds to establishing the links between the input layer and the semantic layer (see figure 32b). In the proposed model, just as

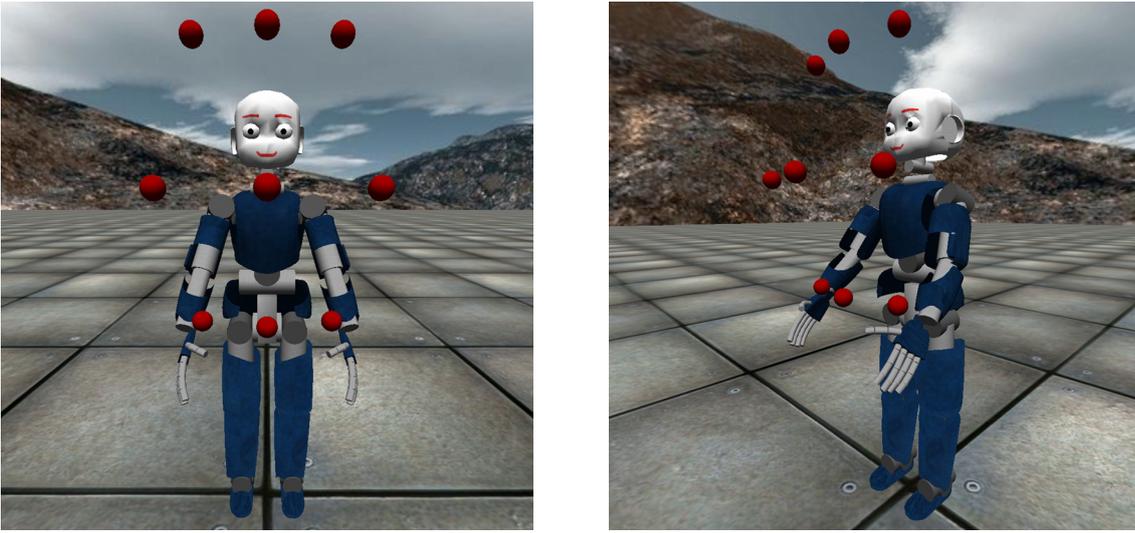


Figure 33: Visualisation of the extent of the iCub’s operational space during motor babbling in the iCub simulation software.

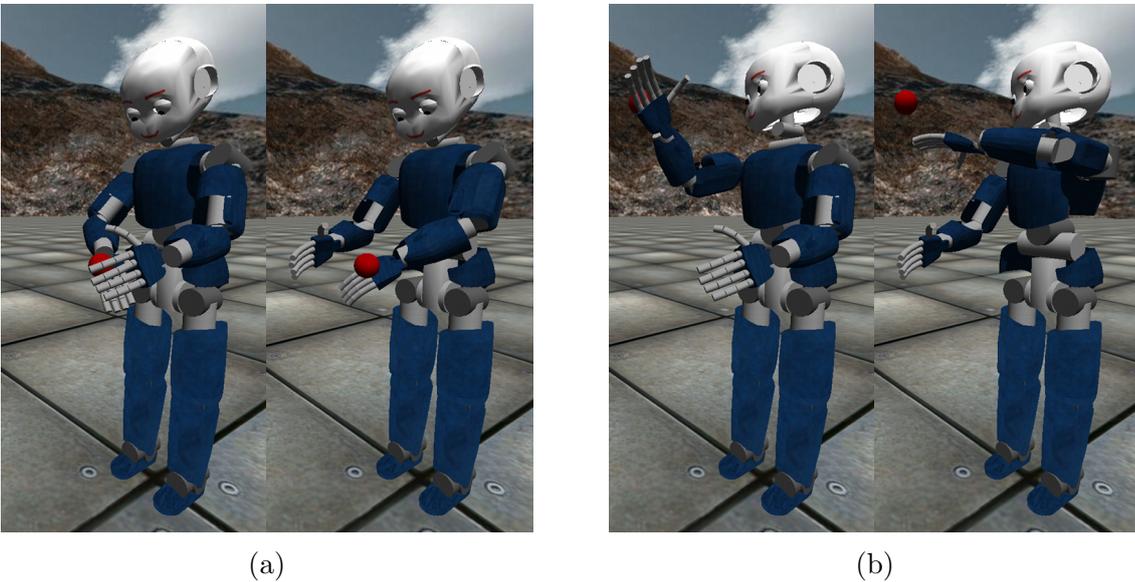


Figure 34: Simulated iCub robot performing motor babbling. The red ball indicates the target position to be fixated at and reached by both arms. Note that the locations in the centre of the robot’s operational space can be reached with both arms (a), while for laterally located positions, due to the robot morphology, only one arm is able to reach the target point (b) .

Map	No. of epochs	Learning coeff.	σ_S^a	σ_F	d_{norm}	Tiling
Gaze	4000	0.006	6.75	0.5	0.192	□
Left arm	24000	0.001	6.75	0.5	0.744	△
Right arm	24000	0.001	6.75	0.5	0.744	△

^a The $\sigma(t)$ parameter of the neighbourhood function decreased linearly starting from σ_S down to σ_F .

Table 4: Parameters of the SOM training algorithm used in the model development process.

in the one of Chen and Verguts (2010), the connection weights between those layers were pre-set manually, implementing place coding with linear scaling and constant variability (see section 7.2.1 and figure 31). It is important to note however, that it has been demonstrated earlier that such a pattern of connections can emerge in a simple, unsupervised training process (Verguts & Fias, 2004).

7.3.3 Learning to Count

From the modelling point of view, this is the crucial part of the proposed development process. When describing their model, Chen and Verguts (2010) express their assumptions about the origin of the spatial-numerical associations as follows:

In the current context we assume that an environmental correlation between symbolic (e.g. Arabic) numbers and physical space (left or right positions) leaves its signature in the brain. [...] Applied to the present context, when children in Western cultures go to school and begin to learn Arabic numbers, Arabic numbers 1–9 are often physically represented from left to right (e.g., on the blackboard and in school books). This practice introduces an environmental correlation between number and space in that smaller numbers (e.g., 1 and 2) tend to occur more often left than right, and larger numbers (e.g., 8 and 9) tend to appear more often on the right. We hypothesize that Hebbian learning processes will pick up this correlation and install a stronger coupling between small numbers and left side of space on the one hand and between large numbers and right side of space on the other (Chen & Verguts, 2010, p. 220).

This assumption was reflected in the model of Chen and Verguts in the way the connection weights in the pathway dealing with the space representation were wired. Importantly however, these weights were in their case pre-set by hand. The model of the acquisition of the spatial-numerical associations proposed in this thesis goes one step further than the work of Chen and Verguts by incorporating the actual spatial biases, similar to the ones described above, into the process of the training of the model. As the result, the proposed model literally picks these correlations up from the environment in which it develops.

The source of the environmental correlations that lead to establishing an association between the small numbers and the left side of space and between the large

numbers and the right side of space, is however slightly different here than that assumed by Chen and Verguts (2010). As reviewed in chapter 2, sections 2.2.3 and 2.3.3, systematic, culture-specific spatial biases appear in the children’s interactions with numbers before they go to school or even start to learn to read (Tversky et al., 1991; Opfer & Furlong, 2011). In the proposed model it is assumed that the children’s tendency to explore sets from left to right (in the ‘Western’ cultures), which is connected with the fact that they also learn to count items in the direction from left to right, is the primary source of these spatial biases.

The process of learning to count is simulated by repeatedly exposing the neural network to an appropriate sequence of numbers, fed to the input layer, while at the same time directing the robot’s gaze, represented by the activations in the gaze map, toward the locations along a row, which extends from the left to the right side of the robot’s operational space. This corresponds to the way children would count items arranged in a row, assuming a consistent, left-to-right spatial bias. The vertical location of the row of items is randomised within the range of the values represented in the gaze map. The resulting co-activations in the input layer and the gaze map are then used to establish the strengths of the links between those two layers, using the Hebbian learning rule (see figure 32c). The data set used to train the model at this stage of the development consists of 50 rows.

7.3.4 Learning Simple Numerical Tasks

In the final stage of the development, the model is trained to perform simple numerical tasks, which are used in the behavioural tests aimed at detecting spatial-numerical associations, i.e. number comparison and parity judgement. This corresponds to establishing the weights of connections between the semantic layer and the decision layer in the What pathway (see figure 32d). My implementation of this process essentially reproduces the methodology described by Verguts et al. (2005). Since the trained neural network is a single layer of linear units, the Widrow-Hoff Delta learning rule (Widrow & Lehr, 1990) is applied. The only required modi-

fication to the training algorithm is connected with the fact that the network is implemented within the firing rate framework. The activation values of the units must therefore be determined after the equations describing the network activity reach a stable state. The training lasts for 30000 epochs, with the learning rate 0.02. Numbers presented to the model in each epoch are drawn randomly from an exponentially biased distribution. This is intended to reflect the differences in the estimated frequencies, with which numbers are experienced in real life (Dehaene & Mehler, 1992). Similarly to Verguts et al., the frequency with which number i is used during the training is proportional to $e^{-0.2i}$.

The remaining weights in the model, namely the weights from the decision layer to the response layer and from the arm maps to the response layer, are set by hand in order to realise various response mappings required for the behavioural experiments aimed at detecting the SNARC effect (e.g. odd number — left and even number — right versus odd number — right and even number — left). The units in the decision layer connect differentially to the response units using the following weight matrices:

$$W_{reg} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (7.3)$$

for the ‘regular’ response mapping and

$$W_{inv} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \quad (7.4)$$

for the ‘inverted’ response mapping. Such a set-up of weights means that the activation values of the response units are proportional to the difference between the activation values of the units of the decision layer. A similar, differential pattern of connections was set-up between every unit of the right and left arm maps and the response units, realising the priming of the response from the Where pathway. The strength of these connections is a parameter of the model, and plays an analogous role to the β_{SNARC} parameter of the model of Chen and Verguts (2010), namely

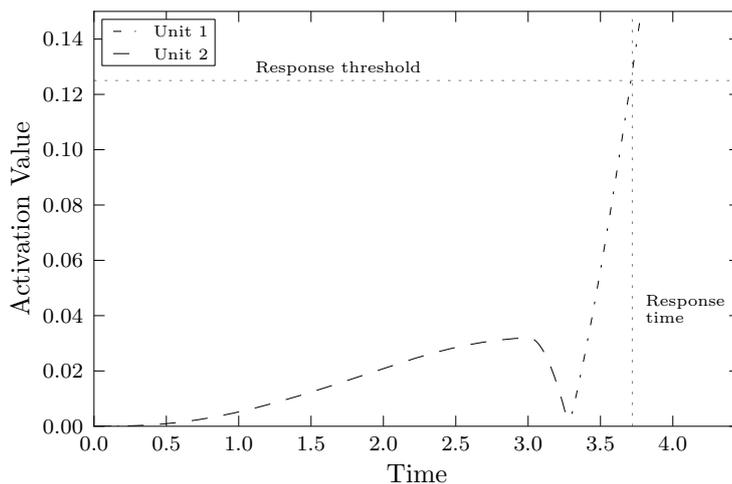


Figure 35: Response time (RT) measurement illustration. The abscissa shows time, expressed in the arbitrary units of the numerical integrator. The ordinate shows the activation value of the two units in the response layer, also expressed in arbitrary units. The response time is determined by the first of the response units, the activity of which exceeds a predefined threshold (here equal to 0.125). In this example, the response is given by unit 1 at $t \approx 3.7$.

it determines the magnitude of the SNARC effect. The used connection weights are reported along with the descriptions of the particular simulations in chapter 8.

7.4 Model Evaluation

All simulation experiments presented in the following chapter involve the measurement of the model's response times (RT). Since the firing rate framework is used, this is implemented by monitoring the activity of the units in the response layer as the integration of the differential equations proceeds. The response is assumed to be given by the model as soon as the activation value of one of the units in the response layer exceeds a threshold value assumed for the task. The unit that exceeds the threshold first determines the response given. If neither response unit exceeds the activity threshold, this is interpreted as a failure to give a response. Such instances are dropped from the RT analysis. The process of the measurement of RT is illustrated in figure 35.

7.5 Neuro-Robotic Model of the Acquisition of Spatial-Numerical Associations — Summary

The model of the acquisition of spatial-numerical associations presented in this chapter is in many aspects similar to the one proposed by Chen and Verguts (2010). It adopts the same multi-pathway architectural design, and the pathway that realises the numerical processing is a reproduction of the corresponding elements of the quoted model (compare figures 30 and 2). It is the spatial representation pathway where the crucial differences between these two models are. Here, this pathway consists of actual spatial maps, which represent the space around the iCub humanoid robot, a physically existing entity. These representations are constructed in a developmental process intended to resemble what happens during the corresponding stage of the ontogeny in humans. Consistent with this, the patterns of weights responsible for the establishment of a link between numbers and space are expected to also come into being as the result of a robotic simulation of the spatial biases in learning to count. In the following chapter it is investigated if the proposed development process leads to a successful reproduction of the behavioural effects that indicate the acquisition of the spatial-numerical associations, what would provide evidence toward a positive answer to the final research question considered in this thesis.

Chapter 8

Simulations of the Acquisition of Spatial-Numerical Associations

In this chapter, the results of the simulation experiments conducted using the model described in chapter 7 are reported. The description is split into four parts, each focusing on different aspects of the proposed model. The first part is concerned with the results of the development process introduced in section 7.3. The remaining three parts describe the simulations that focus on the behavioural effects, in which spatial-numerical associations are manifested, respectively for: the number size and numerical distance affects, the SNARC effect and the Posner-SNARC effect.

8.1 Simulation 1 — Results of the Development Process

8.1.1 Aims of the Experiment

The aim of this simulation is to illustrate the typical results of the development process outlined in section 7.3. This is necessary in order to verify if the chosen values of the parameters of the training algorithms applied at the consecutive stages of the development consistently lead to the desired results.

8.1.2 Procedure

The development process described in section 7.3 was repeated independently 10 times, with randomised stochastic aspects of the training regime (such as the initial weights of the connections in the network and the order of the presentation of the training examples). In all 10 trials, N_I was assumed to be 15, and the same set of proprioceptive data gathered throughout the simulated motor babbling was used. The motor babbling was conducted using the simulated iCub robot. Efforts to transfer the process to the real robot were undertaken, but because of the mounting technical difficulties and the limited time frame of the project, it was not feasible to complete this within the scope of this thesis. All simulations reported in this chapter use therefore the data acquired using the simulated robot.

The essence of the analysis of the results of the model training is the inspection of the neural network components obtained in each of the phases of the development outlined in section 7.3. The obtained networks are assessed, visualised, and compared across trials, what is aimed at providing convincing evidence that the proposed training process has been designed correctly and its outcomes are robust.

8.1.3 Results

As outlined in section 7.3, the first stage of the development of the model involves building the gaze and arm spatial maps and establishing the connections between them. Although there are different quantitative metrics that aid the assessment of the quality of a SOM (cf. section 4.4.3), no single metric is able to capture all the factors involved. As a consequence, it is the visual inspection of the results that turns out to be the most informative in practice, and therefore it is presented prior to the quantitative analysis. In all 10 trials, the results of this stage of training were similar, thus the outcome of only one of the trials is presented below.

The goal of SOM training in general is to map the manifold of the training data in the input space onto the SOM topology. In other words, in a well-trained SOM, the

units close to each other in the SOM topology correspond to nearby areas of the input space. Formally, such a SOM is characterised by a strong positive correlation between the distance between two SOM units in the SOM topology and the Euclidean distance between the weights associated with these units in the SOM input space. Figure 36a shows the relation between these two variables in an instance of a developed gaze map.

A way to visualise the actual mapping of the input space manifold onto the SOM topology is to plot the training data and show the SOM weights and topology on top of it. Such an illustration for the gaze map is shown in figure 36b. The input space of the gaze map is two-dimensional — the first dimension corresponds to the gaze azimuth and the second one to the elevation — which makes the visualisation straightforward, but also enables one to expect that the training of the gaze SOM, the topology of which was assumed to be two-dimensional as well, should not be problematic.

Analogous charts to those in figure 36 for the left arm map obtained in the same training trial are shown in figure 37. Results for the right arm map were similar, due to the symmetry between the robot’s arms. Because in the case of the arm maps the input space is seven-dimensional, the mapping is visualised on three charts (37b, 37c and 37d), with their axes corresponding to the consecutive dimensions of the input space.

Another way to visualise the relation between a developed arm SOM and its input space, is to show which locations of the robot’s arm end effector, in the robot’s operational space, the weights of the SOM units correspond to. This can be done simply by commanding the robot arm to assume the poses defined by the values of the weight vectors of the SOM units and registering the positions of the end effector using the Cartesian controller. The results of this procedure for the left and right arm SOMs obtained in the considered trial are shown in figure 38. The colours of the points in figure 38 encode the locations of the units in the SOM topology (cf. figure 36).

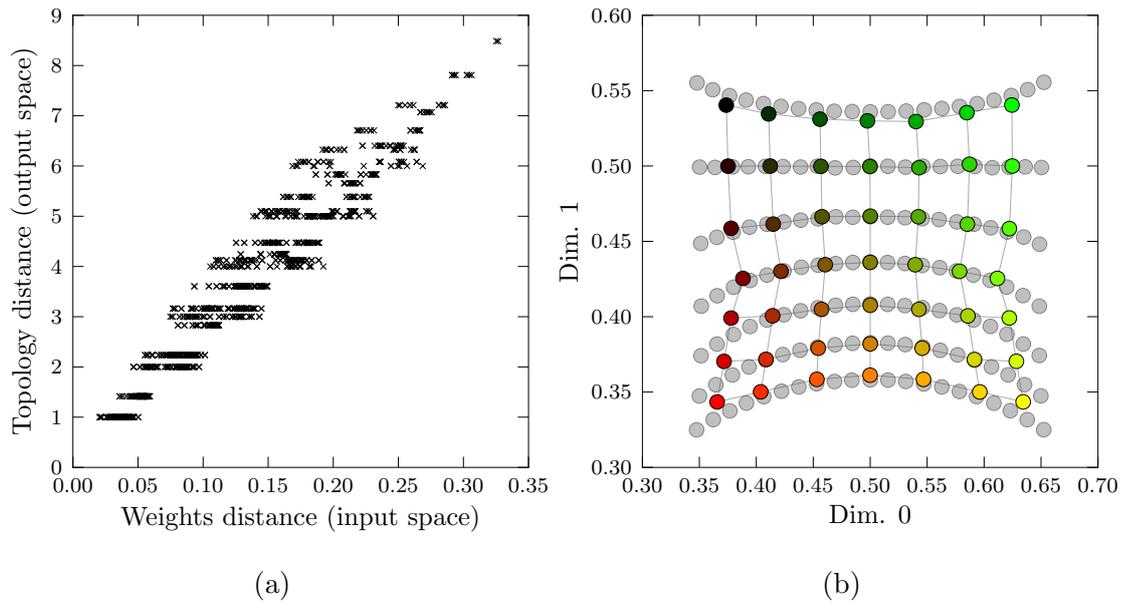


Figure 36: Sample results of the gaze SOM training. (a) distance between SOM units in the SOM output space (i.e. SOM topology) as a function of the Euclidean distance between the weights associated with those units in the SOM input space. (b) SOM weight vectors plotted against the training samples in the input space. The abscissa and ordinate correspond to the input space dimensions (gaze azimuth and elevation angles). Training data points gathered during motor babbling are shown in grey. Coloured points show the weights associated with the SOM units. Colours encode the location of the unit in the SOM topology (here 7×7 with square tiling): the amount of the red component indicates the row, and the amount of the green component indicates the column. Gray lines between the points indicate immediate neighbours in the SOM topology.

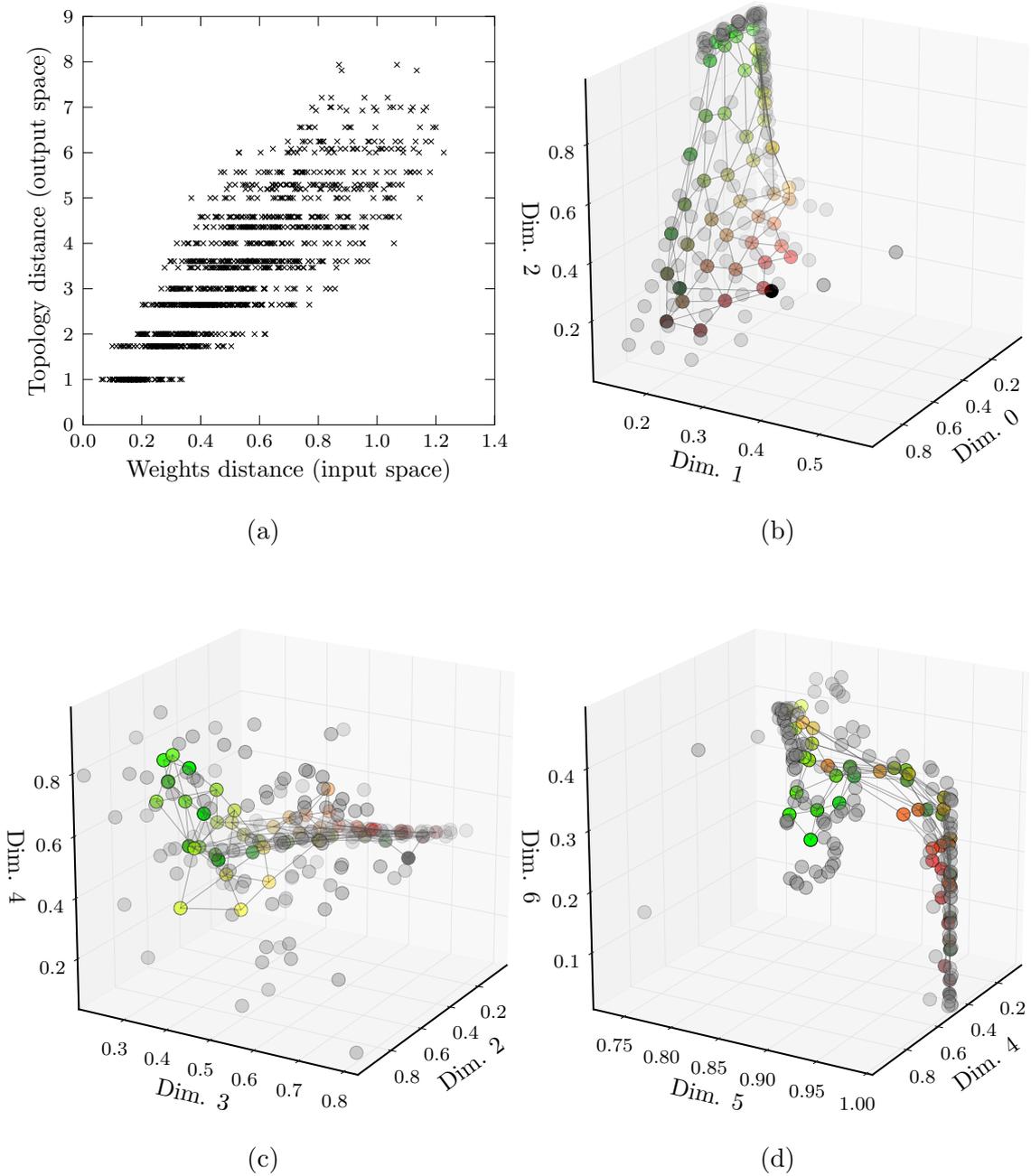


Figure 37: Sample results of the left arm SOM training. The interpretation of the charts is the same as in figure 36. (b), (c) and (d) show the seven-dimensional input space (corresponding to the seven first joints of the robot arm kinematic chain) by presenting 3 consecutive dimensions in each chart.

Finally, the values of the quantitative metrics of SOM quality (the average quantisation error and the topographic error) computed across the 10 trials are shown in table 5.

After the spatial representations in the Where pathway of the model are constructed, the connections between them are established via Hebbian learning. Figure 39 illustrates the obtained patterns of connectivity between the gaze map and the arm maps in the considered training trial. Because the orientation of the nodes in the SOM output space is arbitrary, in different trials the left and right side of the robot’s operational space may be mapped onto different ‘sides’ of the gaze map. For example, the left part of the robot’s visual field may become mapped onto the ‘top’ or ‘right’ part of the gaze SOM. Consistent with this, in different training trials different mappings were obtained. In order to avoid confusion, figure 39 shows a trial, in which the ‘intuitive’ mapping (the one, in which the left side of the robot’s space is mapped onto the ‘left’ side of the visualised map) has been obtained.

In figure 39, the strengths of the connections are shown in the form of small square images. Each small image in figures 39a and 39b corresponds to a unit in the gaze map (thus the image arrays consist of 7×7 images). Note, that the images in the corresponding locations in figures 39a and 39b refer to the same units of the gaze map. The brightness of the pixels in the images indicates the relative strength of the connection between the gaze map unit associated with the image and every unit in the target map (thus each image consists of 7×7 pixels). The brighter the pixel, the stronger the connection, and therefore the brighter the image overall, the more strongly the unit of the gaze map is connected to the appropriate arm map.

Similar Hebbian learning process takes place in the subsequent training stage, which is intended to simulate the systematic spatial biases present in learning to count, hypothesised to lead to the formation of the spatial-numerical associations. Figure 40 shows the relative strengths of the connections between each unit of the input layer and the 49 units of the gaze map, obtained in the considered training trial. Here, each square image corresponds to a unit of the input layer, and therefore,

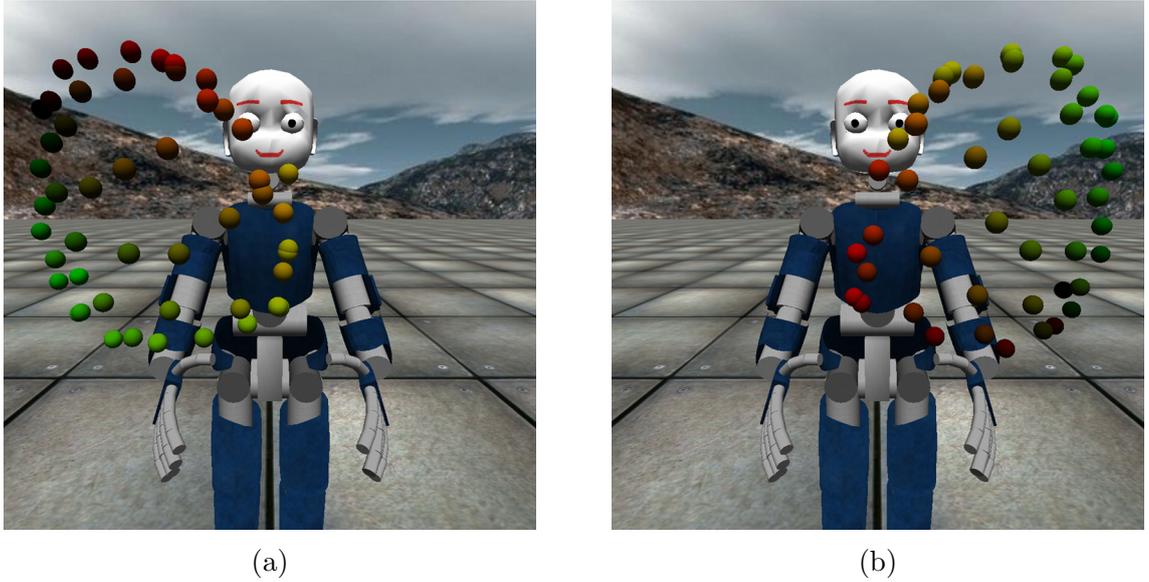


Figure 38: Visualisation of the developed arm maps in the operational space of the robot. Each point shows the position of the robot’s palm when the arm joint angles are equal to a weight vector of a SOM unit: (a) in the right arm SOM, (b) in the left arm SOM. The same method of colour encoding of the SOM topology is used here as in figure 36b.

Metric ^a	Map	Mean	Std. dev.
AQE	gaze	0.012	$6.94 \cdot 10^{-7}$
	left arm	0.161	0.0017
	right arm	0.149	0.0036
TE	gaze	0.0	0.0
	left arm	0.129	0.0527
	right arm	0.092	0.0331

^a see chapter 4 section 4.4.3

Table 5: Descriptive statistics of the SOM quality metrics across the 10 trials of the model development.

to the symbolic representation of the number designated in the figure. Each pixel in the image corresponds to a node in the gaze map. As mentioned earlier, in the presented trial the left side of the gaze map corresponds to the left side of the robot's space.

The final stage of the model development consists of the supervised training of the connection weights between the semantic layer and the decision layer in the What pathway. Three sets of weights are obtained as the result of this process. The first set enables the model to judge the parity of a number. The two remaining sets are needed to realise the number comparison. As discussed in chapter 7, the simulation of the number comparison task requires two semantic layers to be present in the model, and therefore two sets of connections from the semantic layer to the decision layer — one for each copy of the semantic layer — are necessary. The results of the training, across the 10 performed trials, are shown in figure 41.

8.1.4 Discussion

Taken all together, the results of the training of the SOMs in the Where pathway of the model were consistently good, across all 10 trials. The clouds of the points in plots 36a and 37a show clear positive correlation, which confirms that units nearby in the SOM topology represent nearby areas of the input spaces. The cloud in figure 37a is more fuzzy than the one obtained for the gaze map. This should not surprise however, considering the higher dimensionality of the input space in case of the former. Figures 36b, 37b, 37c, and 37d provide evidence that the SOMs span their input spaces quite well, preserving at the same time the structures of the respective map topologies. In case of the arm map, this is particularly evident in figure 37b. The fact that following the training the arm SOMs successfully capture the properties of the training data is illustrated in an elegant way in figure 38. Note, that in this figure the points of similar colours — which correspond to the units close to each other in the SOM topology — are located close to each other in the robot's operational space. This demonstrates correct representation of the

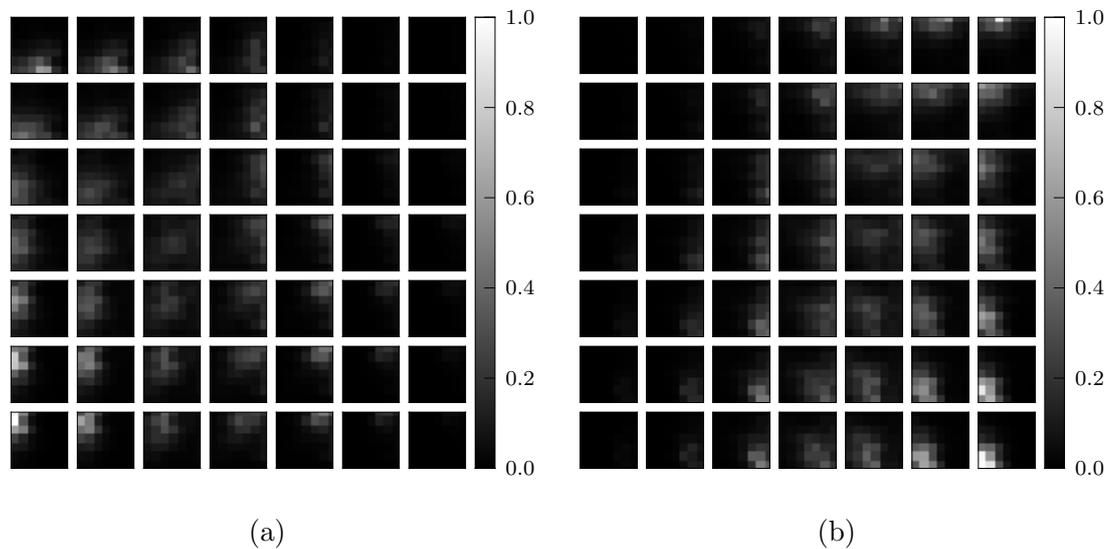


Figure 39: Sample results of the Hebbian learning between the gaze map and the left arm map (a), and between the gaze map and the right arm map (b). Each of the pictures arranged in the 7×7 array corresponds to a gaze map unit, and pixels within each picture correspond to the units of the arm maps. Relative strengths of the connections are shown using the shades of grey.

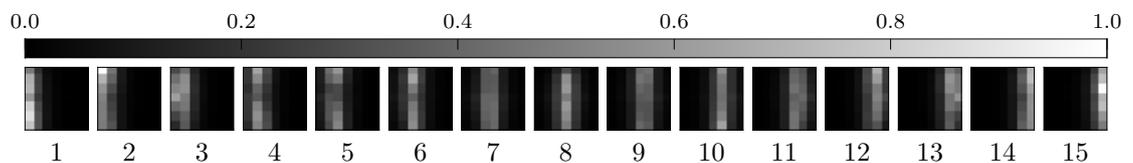
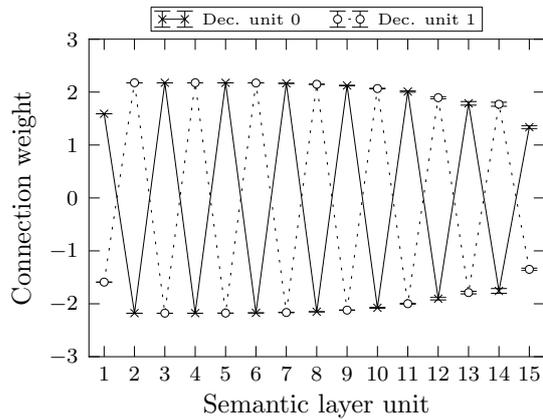
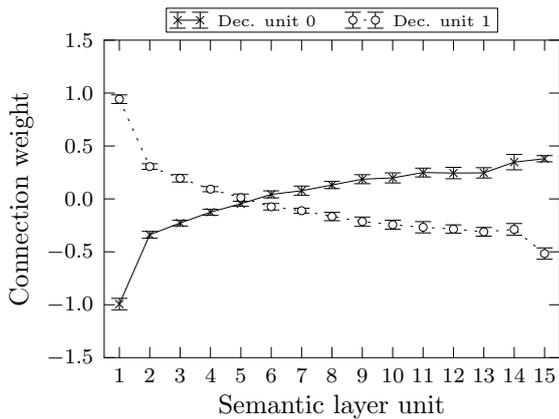


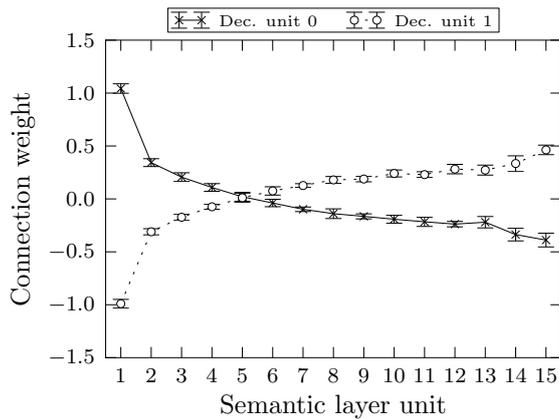
Figure 40: Sample results of the Hebbian learning between the input layer in the What pathway and the gaze map in the Where pathway. Each picture in the array corresponds to one unit of the input layer ($N_I = 15$). Pixels within each picture correspond to the units of the gaze map. Relative strengths of the connections are shown using the shades of grey.



(a)



(b)



(c)

Figure 41: Results of the simple numerical tasks training. The charts show the weights of the connections between the semantic layer and the decision layer for the number parity task (a), and the number comparison task (b) and (c). Error bars show 95% confidence intervals.

robot’s reachable space in the arm maps. Note also that the areas of the operational space represented by the left and right arm maps overlap in the middle, but are, in general, distinct. This fact will be important when the results of the Hebbian learning will be interpreted below. Finally, the values of the SOM quality metrics computed across trials (table 5) provide quantitative evidence of the good quality of the constructed mappings (as indicated by low absolute values of the average quantisation error and of the topographic error), and that the chosen values of the parameters of the SOM training algorithm were robust (as indicated by low standard deviation values across the 10 trials).

The most important feature of the connections established between the spatial maps (figure 39) are the lateral gradients of the overall strengths of the links. In

other words, the left part of the gaze map is more strongly connected to the left arm map, and the right part of the gaze map is more strongly connected to the right arm map (compare figures 39a and 39b). Note that as one moves from left to right along the gaze map, the connections to the left arm map become weaker, and the connections to the right arm map become stronger. This is a crucial pattern of connectivity, which is responsible for the emergence the SNARC effect (Chen & Verguts, 2010). Such a pattern of weights emerges in the proposed model as the result of the fact, that the constructed arm maps represent overlapping, yet distinct parts of the operational space in front of the robot (figure 38). This, in turn, is due to the morphology of the robot, in which some areas of the space around the robot are reachable by one arm but not by the other. As a consequence, after training each of the arm maps ‘over-represents’ its corresponding side of the robot’s peripheral space. Importantly, in the present work the pattern of connections responsible for the SNARC effect is obtained as a result of the robot morphology and the applied training regime, while in the previous works such connections were hand-wired (Gevers, Verguts et al., 2006; Chen & Verguts, 2010).

The second factor necessary to obtain the SNARC effect (and the Posner-SNARC effect) is the actual association of numbers with space. As explained in chapter 7, the aim of the model is to demonstrate that this can happen as the result of systematic spatial biases present in learning to count. The figure 40 shows clearly, that the learning to count simulated using the iCub robot leads to a robust association of the small numbers with the left side of space and the large numbers with the right side of space. As a result, presentation of a number in the input layer of the model will evoke activity in the gaze map, the location of which will depend on the number magnitude.

The above results confirm that the patterns of connectivity which lead to the response time patterns characteristic of spatial-numerical associations consistently emerge in the proposed neural network as the result of the spatial biases present in the artificial set-up intended to resemble children’s learning to count. First, this val-

idates the assumptions of Chen and Verguts (2010) quoted in the previous chapter (see section 7.3.3) about the sources of the spatial-numerical associations in their model, reflected by the way they decided to pre-set the connection weights in the spatial representation modules of their neural network. Second, and more importantly, this constitutes computational evidence which suggest a positive answer to the research question 4, by demonstrating that the factors considered in the present experiment when designing the modelled learning to count scenario are sufficient to account for the appearance of the connection patterns which cause spatial-numerical associations. For the time being, this evidence is of course partial, as it is based only on the visual inspection of the obtained neural network weights. Further evidence, based on the investigation of the behaviour of the model in simulated simple numerical task, will be sought in the subsequent simulations.

Finally, it is appropriate to comment on the results of the training of simple numerical tasks. As indicated by the low error bars in figure 41, the obtained weights of the connections from the semantic layer to the decision layer were quite consistent across trials. All three charts in this figure show a signature of the distribution of the frequency of number presentation during training being skewed toward small numbers (cf. section 7.3.4). In figure 41a, the values of the weights decrease, and the variability across the trials increases as the numbers increase. In figures 41b and 41c, the patterns of the weights have a monotonic, but compressive character, with the crossover point biased toward smaller numbers, rather than lying in the middle of the considered interval. The latter is especially important, because it is responsible for the presence of the number size and numerical distance effects in the number comparison task (Verguts et al., 2005, p. 77). The dip in the absolute value of the weights for number 1 in figure 41a is most likely a ‘border effect’, connected with the fact, that the total activation induced in the semantic layer for the extreme numbers (1 and 15) is smaller than for the other numbers. Overall, the obtained results are consistent with the findings of Verguts et al. (2005).

8.2 Simulation 2 — Number Size and Numerical Distance Effects

8.2.1 Aims of the Experiment

Number size and numerical distance effects are among the most ubiquitous findings in experimental studies involving numerical tasks (see chapter 2). Despite the fact they are not directly connected with spatial-numerical associations, they can still be considered as a fundamental benchmark behaviour for cognitive models in mathematical cognition. The purpose of the present simulation is therefore to investigate if the proposed model exhibits the number size and numerical distance effects in the number comparison task.

8.2.2 Procedure

In the present simulation (and in all subsequent experiments described in this chapter), the 10 instances of the model developed in the 10 trials conducted for the purposes of simulation 1 were used. Number comparison task was simulated using the model in configuration shown in figure 42. In order to simulate short-term memory (what is necessary, since the simulated numerical task requires two operands), two copies of the input layer and of the semantic layer were included in the What pathway of the model (following the practice of Verguts et al., 2005). The decision layer was connected with the response layer using the weight matrix provided in equation 7.3. The left-sided response was to be given when the first number was larger than the second one, and the right-sided response otherwise. The units of the arm SOMs were connected to the corresponding-side unit of the response layer with connection weights equal to 25, and to the opposite-side unit -25 . The value of the response threshold was set to 0.125.

Following the design of the corresponding experimental paradigms (see for instance Schwarz & Stein, 1998), the response times of the model were measured for

all pairs of distinct numbers from 1 to 15, in both possible orders (smaller number first and larger number first). This means that each model was tested on 210 number pairs. Latencies obtained for both left- and right-sided responses were then aggregated across the 10 trials. The instances when the response given by the model was incorrect, or when it was not given at all, were not included in the analysis.

8.2.3 Results

The response times (RT), aggregated over the 10 tested instances of the model, are shown in figure 43. The figure 43a shows RTs for the left-sided response (first number larger), and the figure 43b for the right-sided response (second number larger). In both charts, the abscissa shows the smaller of the two numbers being compared, and the data series correspond to the constant numerical distance between the two operands. For conciseness, the charts do not include all performed comparisons. Outside of the number range presented in the figures, the results became less and less clear, what was caused by the increasing number of incorrect or missing responses. The overall accuracy of the responses of the 10 models was 92.3% (the response was incorrect or missing in 162 out of 2100 cases).

8.2.4 Discussion

In the context of the number comparison task, the behavioural effects considered in the present simulation mean that it is more difficult to compare larger numbers than smaller numbers (the number size effect) and that the numbers that are close to each other are more difficult to compare than the numbers which are further apart (the numerical distance effect). The charts in figure 43 show clear evidence of both effects being exhibited by the model in the presented range of numbers. The numerical distance effect is manifested in that the data series are located the higher on the chart, the lower the numerical distance, to which they correspond. The number size effect in turn, is evident in the data series rising along with the

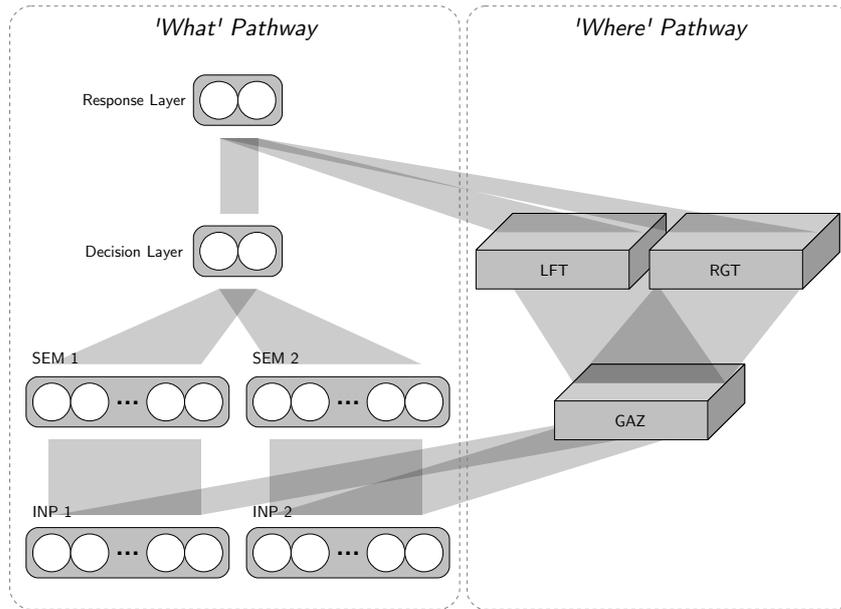


Figure 42: Configuration of the model used in the simulations of the number comparison task.

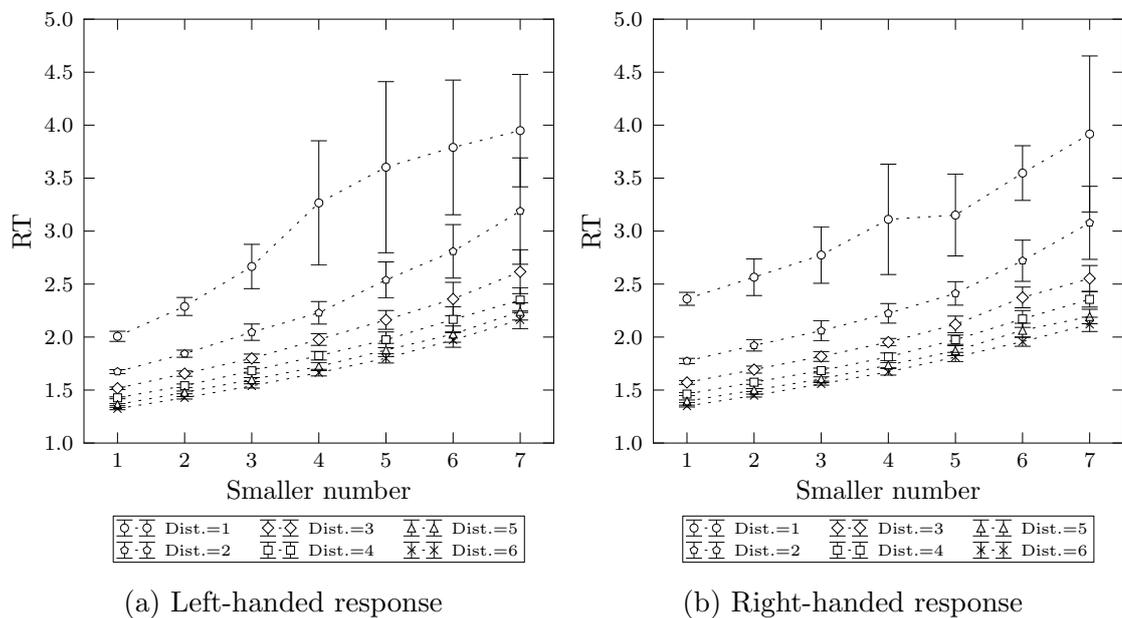


Figure 43: Simulated response times of the model in the number comparison task for the left-handed response (a) and the right-handed response (b). The ordinate shows the RT in the arbitrary time units of the numerical integrator. The abscissa shows the smaller of the two operands in the comparison. Data series correspond to the numerical distance between the comparison operands. Error bars show 95% confidence intervals.

increasing magnitude of the operands (i.e. along the abscissa). The results of the simulation confirm therefore, that the proposed model exhibits the number size and numerical distance effects in the number comparison task.

The considered effects are present in the behaviour of the model as the consequence of the patterns of weights obtained between the semantic layers and the decision layer. Because the strengths of these connections change monotonically with the increasing numbers, the greater the distance between the numbers being compared, the stronger the activation of the units in the decision layer, and, hence, the shorter the response time. The number size effect is the result of the fact that the weights change (increase or decrease, respectively) in a compressive manner (see figure 41). For a fixed distance between the numbers, the resulting activation of the decision layer units is the smaller (and consequently the response time the longer), the larger is the absolute magnitude of the numbers (see Verguts et al., 2005).

8.3 Simulation 3 — The SNARC Effect

8.3.1 Aims of the Experiment

The present simulation aims at demonstrating that following the proposed development process, the model exhibits the well-established effect believed to be a manifestation of the spatial-numerical associations, namely the SNARC effect (see chapter 2 section 2.2.3). This is one of the crucial simulations, since the goal of the present modelling effort is to demonstrate that the spatial biases present in counting may lead to the acquisitions of the spatial-numerical associations.

8.3.2 Procedure

As reviewed in chapter 2, the SNARC effect emerges in various numerical tasks. Two of those tasks, that the proposed model is capable of simulating, are number comparison and parity judgement. The SNARC effect is obtained by measuring the

response times of the subject in two reverse response mappings (e.g. press left when the number is odd versus press right when the number is odd), and calculating the difference in the RTs for each number between the right-handed and the left-handed response. The ‘Western’, left-to-right SNARC effect is then found as a downward trend in the difference of the RTs for the increasing numbers, meaning that the small numbers are responded to faster with the left hand than with the right hand, and, conversely, that the large numbers are responded to faster with the right hand than with the left (Fias et al., 1996).

In the simulation of the number comparison task, the configuration of the model (including the weights of the connections between the arm maps and the response layer as well as the response threshold) was the same as the one used in simulation 2 (figure 42). The model configuration employed to simulate the parity task is shown in figure 44. The weights of the connections between the arm maps and the response layer for the parity task were set to 50 (the corresponding response unit) and -50 (the opposite unit). The double-fold increase in the connection weights with respect to the number comparison task ensures a comparable magnitude of the SNARC effect in both tasks, as it compensates for the fact that in the parity task only one input layer is present in the model, and therefore the activations propagated in the Where pathway are approximately half as strong as in the comparison task. The response threshold used in the parity task was set to 0.8. In both tasks, the two reverse response mappings were realised by applying the weight matrices provided in equations 7.3 and 7.4 between the decision layer and the response layer.

In the number parity task, the models were tested on all 15 numbers. The number comparison was made between numbers 1–7 and 9–15 and a fixed comparison standard equal to 8, in both smaller-number-first and larger-number-first arrangements (see Gevers, Verguts et al., 2006). In both tasks, every stimulus was presented twice, once for the ‘regular’ response mapping and once for the ‘inverse’ one. The differences in the RTs were measured and aggregated across the 10 copies of the model obtained in simulation 1. Incorrect or missing responses were not included in

the RT analysis.

8.3.3 Results

The calculated differences in the response times between the right- and left-handed responses for every number in both tasks are shown in figure 45. Decreasing or increasing trend in such a chart indicates the presence of the SNARC effect (Fias et al., 1996). The models achieved 100% correct responses in the parity task and 88.7% in the number comparison task.

8.3.4 Discussion

The negative slope of the difference between the right- and left-handed response times for increasing numbers is clearly visible in case of the the parity task (figure 45a), indicating a robust SNARC effect. The effect is also present in the comparison task (figure 45b), although in this case the RT pattern has a slightly different, step-like shape. For the target numbers smaller than the comparison standard (i.e. for numbers 1–7), the RT difference is consistently greater than 0, what means that for these numbers the left-handed responses were faster than the right-handed ones. The opposite holds for the numbers greater than the comparison standard (numbers 9–15). The non-monotonic shape of the effect in the number comparison task is a consequence of an interaction between the SNARC effect and the numerical distance effect. This is consistent with the results of the behavioural studies as well as of the modelling experiments by Gevers, Verguts et al. (2006) and Chen and Verguts (2010). Gevers, Verguts et al. explain this phenomenon as follows (note though, that in their simulations the comparison standard was equal to 5):

The model assumes that the categorical effect is the result of an interaction between the distance effect and the time course of the SNARC effect. Recall that the SNARC effect becomes stronger with increasing time. Taken together with the fact that the slowest latencies are those numbers closest to the standard (e.g., distance effect), a categorical shape results. More specifically, because of the distance effect, the latencies to

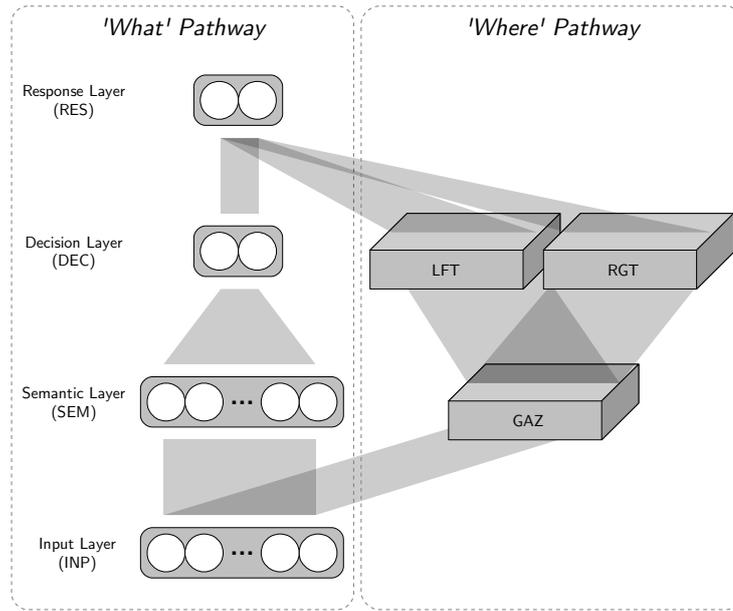


Figure 44: Configuration of the model used in the simulations of the number parity task.

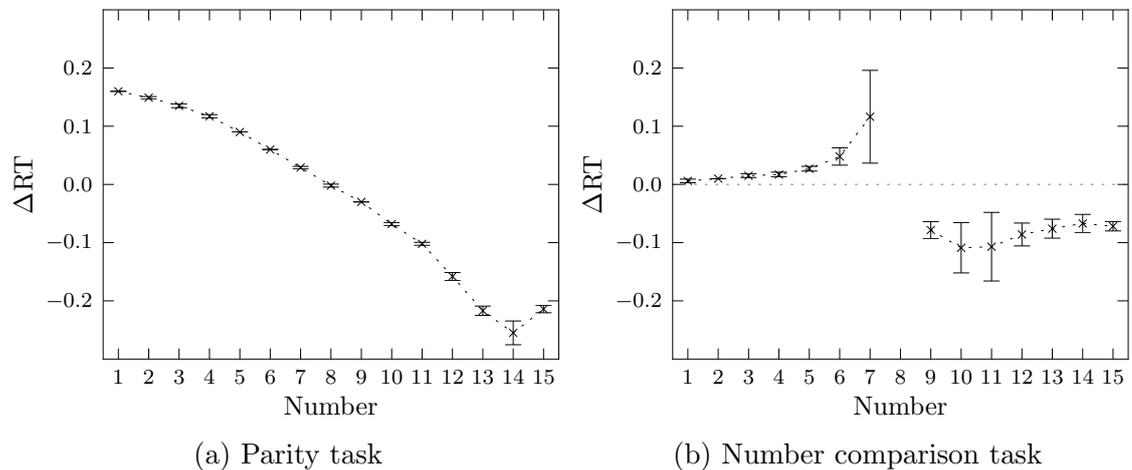


Figure 45: Simulation of the SNARC effect in the parity judgement task (a) and the number comparison task (b). The abscissa shows the target number (in case of the comparison task, the target number is to be compared against a fixed standard 8). The ordinate shows the difference between the right-handed RT and the left-handed RT. Error bars show 95% confidence intervals.

the numbers 4 and 6 will be longer than to the numbers 3 and 7. Because the SNARC effect is strongest with the slowest latencies, the number 4 and 6 will be influenced more by the SNARC effect than the numbers 3 and 7. This effect breaks the continuity of the SNARC curve, pushing the dRT value for number 4 up (stronger effect) and pushing the dRT value for number 6 down (stronger effect). As a result, the observed dRT shape becomes categorical rather than continuous. (Gevers, Verguts et al., 2006, p. 40)

One unexpected feature of the SNARC effect obtained in case of the parity task is the non-monotonic behaviour of the SNARC curve for number 15 (see figure 45a). Narrow confidence intervals suggest this is not a mere artefact caused by large variability across the models, but a consistent result. The reason behind this is most likely the border effect obtained in the training of the neural network performing the parity task (cf. section 8.1.4 and figure 41a). It results from the fact that the input and semantic layers consist of a small, finite number of units. It would be possible to eliminate this artefact by compensating for the border effect. This is possible, for instance, by normalising the weights of the connections between the input layer and the semantic layer in such a way that the numbers ‘located’ near the edges of the semantic layer are not represented with a lower overall level of activity of the units in the layer.

The results of the experiment confirm that, as predicted in section 8.1.4, the connectivity patterns obtained in the model as the consequence of the proposed development process enable it to exhibit the SNARC effect. The mechanism by which this happens is the following. First, the association of numbers with space is obtained in such a way that small numbers are linked with the left side of space and the large numbers with the right side (Hebbian links between the input layer and the gaze map). Moreover, the left part of the gaze map is more strongly connected to the left arm map, and the right part of the gaze map is more strongly connected to the right arm map (as the consequence of the robot’s morphology). Such a connectivity pattern causes that, for instance, when a small number is presented at the input to the model, an activity is evoked in the left side of the gaze map, which is then propagated strongly to the left arm map, but weakly to the right arm map. This

results in a stronger level of activity in the left arm map than in the right arm map, what, via the connections between the arm maps and the response units, facilitates the response with the left hand. If the activity arriving from the Where pathway is congruent with the decision made in the What pathway, the response time of the model is shortened. In turn, when the activity in the Where pathway is incongruent with the task answer, more time is necessary for the response units to resolve this ‘conflict’. The net result of this is the SNARC effect. The origins of the SNARC effect in the proposed model are in agreement with the findings of brain imaging studies, which provide the evidence for a dip toward the incorrect response preparation in the incongruent conditions, indicated by the lateralised readiness potentials recorded at the motor cortex (Gevers, Ratinckx, De Baene & Fias, 2006; Chen & Verguts, 2010).

8.4 Simulation 4 — The Posner-SNARC Effect

8.4.1 Aims of the Experiment

In addition to the SNARC effect, the associations between numbers and space are found in the experiments utilising the attention cuing paradigm. Numbers presented as cues involuntarily affect the times necessary to detect visual targets, which appear either in the left or in the right side of the visual field of the subject (Fischer et al., 2003). This effect, called the Posner-SNARC effect, suggests that numbers cause shifts of attention toward the side of the visual field they are associated with. The goal of this simulation is to investigate if the Posner-SNARC effect is present in the behaviour of the proposed model.

8.4.2 Procedure

The model configuration used to simulate the Posner-SNARC effect is shown in figure 46. Since the experimental paradigm does not involve any numerical tasks, but only

visual target detection, most of the components of the model are irrelevant here and thus have been removed. Similarly to the way the Posner-SNARC effect was simulated by Chen and Verguts (2010), the read out of the response in this task is taken from the spatial representation layer (here, the gaze map). The two sides of the gaze map are connected to the response units in such a way, that 24 ‘left-most’ units in the map connect with the weight 1 to the left response unit, and with the weight -1 to the right one. The reverse weights are set-up between the response units and the 24 ‘right-most’ units of the gaze map. One gaze map unit remains unconnected, in order not to unfairly bias the response times toward one side.

The experimental design in this simulation followed that of Fischer et al. (2003) and Chen and Verguts (2010). As a prime, either number 1 or 15 was presented, for a duration of 2 time units of the numerical integration. This was followed by a variable delay, during which nothing was presented to the model (which lasted for 0.2, 1.0, 1.8, or 2.6 time units). After the delay, the visual target was shown. The latter was simulated by inducing in the gaze SOM a pattern of activity corresponding to the presentation of an input vector representing an extremely left or right spatial location within the boundaries of the space represented by the map. The numbers presented as cues did not predict where the target will appear, that is each of the numbers was tested with both left and right locations. The response threshold in the task was set to 0.125. Since the task is simply to detect the target, it did not matter with which response unit the answer was given.

8.4.3 Results

The response times obtained in the Posner-SNARC effect simulation using the proposed model are shown in figure 47.

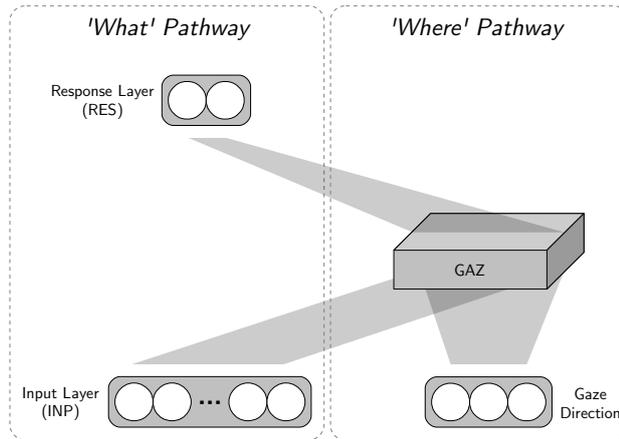


Figure 46: Configuration of the model used in the simulations of the visual target detection task.

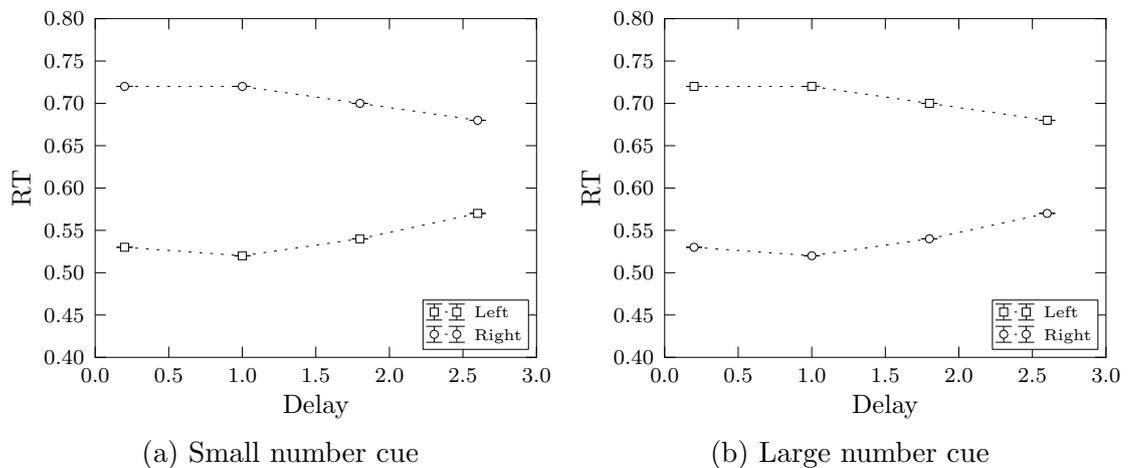


Figure 47: Simulation of the Posner-SNARC effect in the visual target detection task. The charts show the times needed to detect targets that appear in the left or right side of the visual field of the robot. Chart (a) shows response times when a small number is shown as a cue. Chart (b) shows response times for a large number. The abscissa shows the delay between the priming stimulus (the number) and the visual target, expressed in the arbitrary time units of the numerical integrator. The ordinate shows the response time, in the same units as the abscissa. Error bars show 95% confidence intervals.

8.4.4 Discussion

Figure 47 clearly indicates that the signature feature of the Posner-SNARC effect is exhibited by the proposed model. When a small number is shown as a cue, the visual target is detected faster if it appears on the left side than when it appears on the right side (figure 47a), and, conversely, when a large number is shown as a cue, the visual target is detected faster if it appears on the right side than when it appears on the left side (figure 47a).

The effect is caused by the fact that the numbers are associated with space (Hebbian links between the input layer and the gaze map), so a small number presented at input evokes some activity in the left side of the gaze map, priming at the same time the left response unit. If the visual target that appears afterwards is located in the same side of the visual space, the detection time is shortened as a result; if it appears in the opposite side of the visual space, the response layer has to overcome the primed activity first, and the response time is prolonged. Note that the longer the delay between the prime and the stimulus, the weaker the magnitude of the effect. This is to be expected, as the primed activity in the response layer fades out with time due to the self-inhibition present in the neural dynamics.

A notable feature of the obtained results is virtually non-existent between-trial variability (see figure 47). This is caused by the robustness of the results of the gaze SOM development, in which very good quality of the mapping was consistently obtained (cf. simulation 1). This is not surprising, since the construction of a two-dimensional map of a quite uniform two-dimensional input space is not particularly challenging for the SOM training algorithm.

8.5 Summary

Throughout the past two chapters I described a neuro-robotic model of the acquisition of the spatial-numerical associations, and simulated a variety of behavioural experiments using it. The simulation experiments demonstrated the validity of the

proposed model and of the development process associated with it, in the sense that they were shown to lead to the reproduction, at least in the qualitative sense, of the most important behavioural effects in the context of the simulated numerical tasks. Namely, the response times of the model showed the signatures typical for the number size and numerical distance effects in the number comparison task. Furthermore, the SNARC effect in the same task, as well as in the parity judgement task was obtained, showing that the proposed developmental process indeed leads to the establishment of the spatial-numerical associations. This conclusion was reinforced by the manifestation of these associations also in a non-numerical visual target detection task.

Importantly, the obtained results can be considered an evidence, obtained through computational modelling, that spatial-numerical associations can arise as the result of systematic spatial biases present when children learn to count, which are connected with the culturally-specific tendency of the children to enumerate sets of items either from left-to-right in the ‘Western’ cultures, or from right-to-left in the ‘Middle-Eastern’ cultures (Tversky et al., 1991). To the extent the proposed neural network and the associated development process capture the essential factors involved in the real-life situation being modelled, this suggest that the answer to the fourth and final research question posed at the beginning of this thesis is positive.

Chapter 9

Conclusions

In this final chapter of the thesis its most important results are reviewed and the outcomes of the conducted experiments are critically discussed, in order to highlight the areas which require future work. The models and simulations described in parts II and III of the thesis were aimed at answering four research questions posed in section 1.2. These questions are repeated below, along with the short summaries of the corresponding results and the answers which follow from them.

Research question 1: How does mastering the count list prior to learning to count within the respective range of collection sizes affect the subsequent process of learning to count? The results of the simulation experiment described in chapter 6, section 6.2 have shown that, when the training of the model of the contribution of gestures to learning to count introduced in chapter 5 is explicitly divided into two stages — the acquisition of a part of the count list, and learning to count the collections of items within the range of this count list — the model consistently achieves significantly better counting accuracy within the equivalent amount of training (in terms of the number of the performed updates of the model’s weights) than when it is trained to recite the number words and to count at the same time. This suggests that being equipped with a sufficiently long count list ahead of learning to count within its range may enable one to complete the latter process faster.

Research question 2: Can counting gestures, represented in the form of the values of arm joint angles that change over time, contribute toward the improvement of

the counting accuracy? In the experiment described in chapter 6, section 6.3, the counting performance of the model from chapter 5 trained to count in the presence of the proprioceptive signal formed from the values of the joint angles of the arm of the humanoid robot iCub performing the counting gestures turned out to be consistently higher than when the model was trained to count based on the visual input only. Therefore, even a relatively simple machine learning system is able to exploit counting gestures modelled in this way to improve its counting accuracy. This constitutes computational evidence that the proprioceptive information connected with the counting gestures are indeed a salient embodied cue in the context of learning to count, already in their raw, unprocessed form. Further investigation of the performance of the model trained to count based on the counting gestures alone revealed that, in the case of the proposed neural network, the proprioceptive signal facilitated the extraction of information from the visual modality.

Research question 3: Is the spatial correspondence between the items being enumerated and the indicating act performed during counting an important characteristic of the counting gestures? In the simulation reported in chapter 6, section 6.4, ‘natural’, spatio-temporal counting gestures, in which the arm postures correspond 1-to-1 to the spatial locations of the items being counted, have been contrasted with ‘rhythmic’ gestures, constructed in such a way that, although they still corresponded to the correct counting in the temporal sense, there was no 1-to-1 correspondence between the arm postures and the locations of the items. The results have shown that such rhythmic gestures allowed the model to obtain even higher counting accuracy than the spatio-temporal counting gestures, however, in the case of the former, the neural network actually completely disregarded the information in the visual modality (as revealed by the failure to find a statistically significant difference in the counting accuracy for the experimental conditions with and without the visual input for the rhythmic gestures). In other words, when the gestures not characterised by the spatial correspondence to the counted items were supplied, the neural network ‘counted’ the gestures themselves, rather than the items in the visual

input. This constitutes computational evidence that the aforementioned spatial correspondence is a crucial property of the counting gestures which enables them to fulfil a facilitating role in the context of counting.

Research question 4: Can consistent spatial biases present in children's learning to count be a source of the spatial-numerical associations later in life? In chapter 7, the computational model of the interactions between numbers and space published by Chen and Verguts (2010) has been extended with rich embodied representations of space linked to the simulated body of the humanoid robot iCub. In addition, a developmental learning process for the model has been proposed, intended to reflect the progress of the development of children's numerical knowledge in the early years of life. Linking the neural network with a robotic body and thus embedding it in a developmental learning environment made it possible to incorporate in the development process of the model spatial biases resembling those present in the children's learning to count. The simulations described in chapter 8 have shown that these biases indeed lead to the emergence of the patterns of connectivity inside the model which allow it to exhibit the behaviours typical for the spatial-numerical associations, as well as demonstrated the presence of these behaviours in the response time patterns of the model. This constitutes computational evidence that children's culturally-dependent tendency to explore sets of items consistently in a specific lateral direction (left-to-right or right-to-left) is sufficient to account for the establishment of the lateral spatial-numerical associations later in life. Importantly, this finding is consistent with the latest results from the psychological studies with pre-reading children (Opfer & Furlong, 2011), and in opposition to the views about the sources of the spatial-numerical associations which have been dominant only until recently (see Göbel et al., 2011, for review). Since the embodied model of the acquisition of the spatial-numerical associations through learning to count described in chapter 7 has been formulated, implemented and published (Ruciński, Cangelosi & Belpaeme, 2011) in parallel to the hypothesis about the importance of the counting routine in the context of spatial-numerical associations being explicitly

put forward by Opfer and Furlong (2011, p. 691), obtained results can be considered an instance of a prediction resulting from the simulations, which has subsequently found confirmation in the behavioural data.

At this point it is important to stress the crucial role of employing the cognitive developmental robotics approach, and of incorporating in the modelling process of the artificial humanoid body of the iCub robot, for the results obtained in the context of the research questions 2–4. The experiments described in chapter 6 (where research questions 2 and 3 were addressed), employed the kinematic chain of the right arm of the iCub robot to construct a realistic embodied representation of the counting gestures (cf. section 5.2.2). This representation is one of the crucial factors which distinguish the proposed model from the previous efforts in a similar direction. As reviewed in chapter 3, although multiple models of learning to count exist to date, none of them has addressed the issue of the contribution of gestures, and only one included gestures at all. This results from the fact that modelling of the counting gestures, which are by nature an embodied phenomenon, poses a significant difficulty when tackled with purely computational means. Formulating a model of the contribution of gestures to learning to count requires the researcher to make assumptions about the representation of the bodily contribution and of the environment, which, in the case of modelling based on disembodied computations, will always be arbitrary to some extent. Adopting the developmental cognitive robotics approach in the present work alleviated this problem in an elegant way. The artificial robotic body of iCub provided the bodily contribution which can be considered much less arbitrary, because it comes from a physically-existing, humanoid body, which interacts with a real-world environment. Although the need for assumptions regarding the representations has not been removed entirely — they are in fact implicit in the choice of the robotic platform — it is clear that, in the context of the embodied phenomena, developmental cognitive robotics has a considerable advantage over approaches based solely on computation.

Note also, that the extent, to which the answers to the research questions 2 and 3

provided herein in the context of an experimental robotic set-up can be considered relevant to the investigation of human cognition depends crucially on the degree to which the proposed representation of the counting gestures corresponds to the actual proprioceptive information available in the human body. In the light of this, it becomes especially important that the employed robotic platform, iCub, has been designed with a lot of emphasis put on the capabilities connected with manual manipulation (cf. chapter 4, section 4.3). Although whether the artificial proprioception of the iCub can be considered corresponding to a sufficient degree to that of humans is open for debate, as it has been argued in section 5.2.2, the approach to represent gestures based on the pointing performed by a humanoid robot is no doubt a significant step forward in comparison to the abstract and disembodied representations of the indicative acts employed in the context of the modelling of learning to count to date (Ahmad et al., 2002).

The adoption of the developmental cognitive robotics paradigm is also the crucial factor differentiating my modelling methodology from that of the authors of the previous models in the context of the efforts to investigate of the ontogeny of spatial-numerical associations (research question 4), described in chapters 7 and 8. In addition to proposing a neural network capable of exhibiting the considered behavioural effects, I put forward a development process for this network (which, as reviewed in chapter 4.2, is one of the fundamental aspects of the developmental cognitive robotics methodology), designed in a way that is consistent with the available data about the acquisition of the numerical knowledge by children. This enabled me to *demonstrate* that certain crucial patterns of connectivity may emerge in the model in the course of the development, in contrast to the authors of the earlier works, who had to *assume* this is the case.

Furthermore, also in this case, involving the artificial body of iCub made it possible to reduce the level of arbitrariness of the assumptions made in the model design, and, more importantly, to provide an embodied explanation of the observed phenomena. As an illustration of the latter point, it is worth to consider the spatial

gradients in the space representation module of the model of Chen and Verguts (2010, see section 3.4). Therein, these gradients were set-up by hand-wiring the desired patterns of connections. In the present thesis, similar patterns were shown to emerge in the course of the process of development, as the result of the morphology of the simulated iCub robot. This robot has one head and two arms, what justifies including three separate spatial maps for each of these body parts in the model. The arms of the robot can reach overlapping, yet distinct areas in space, what leads to obtaining the aforementioned connectivity patterns in the neural network. Finally, the spatial representations included in the model are, in contrast to the model of Chen and Verguts, *real* spatial maps, in which activations correspond to specific positions of the actual, humanoid limbs of the robot. This again illustrates how adopting the cognitive robotics approach to the investigation of the phenomena which have a strongly embodied character, helps to reduce the arbitrariness of the resulting model, and how cognitive developmental robotics can supplement in an elegant way computational modelling.

Having reviewed the successes of the thesis, I am now going to consider the areas in which the results of the experiments turned out to be imperfect or altogether incorrect, as these point to the directions where future work in the context of the considered research questions is needed.

9.1 Future Work — Research Questions 1–3

Since the majority of the issues with the results of the simulation experiments which addressed the research questions 1–3 is connected with the lack of quantitative fit to the human data, it is appropriate to start the discussion by comparing the experimental protocols employed in both cases, highlighting the similarities and differences between them, and explaining where the latter come from.

9.1.1 Comparison of the Experimental Protocols

As mentioned several times throughout the present thesis, the simulation experiments described in chapter 6 have been modelled after the experimental work of Alibali and DiRusso (1999), with the direct comparison of the obtained counting accuracy figures and of the counting error patterns in mind. The similarities and differences between the two experimental designs are best illustrated by comparing figures 6 and 15, which can be found on pages 124, and 155, respectively. As it has been explained in chapter 5, the young subjects of the psychological experiment correspond to the different instantiations of the proposed model in the simulations. The artificial neural networks, just as children, are characterised by between-subject variability, realised through the randomisation of the initial connection weights, as well as through the stochasticity of the applied training algorithms. Both experiments required therefore aggregation and statistical analysis of the data across a sample of several subjects. Furthermore, both experiments had in place certain exclusion criteria, connected primarily with the baseline counting competence required from the subjects (a least 15% counting accuracy in the case of children, and the ability to recite the count list up to 10 for the neural networks).

The majority of the differences between the two considered experimental protocols is connected with the fact that simple recurrent artificial neural networks are deterministic once trained, that is they produce the exact same behaviour when presented with the exact same input pattern (of course assuming the same initial internal state of the network). Therefore, in the case of the model, there is no *within*-subject variability for a fixed stimuli. The situation is completely different in the case of children, who undergo constant long-term development and whose performance is potentially affected by short-term effects of the performed tasks. A psychological experiment must take this into account and be designed in such a way that the influence of the task sequence is minimised, or at least possible to detect. For example, in the case of the experiments of Alibali and DiRusso (1999), as indicated in figure 6, the experimental condition aimed at investigating the spontaneously

performed counting gestures had to take place as the very first one (since subsequent conditions included detailed instructions regarding the gestures, once these were undertaken, children's pointing could not be regarded as spontaneous any more), and the 'puppet incorrect' condition must have been tested as the very last one (in order not to undermine the children's belief in the puppet's counting competence, which was important in the other conditions). However, the order of the remaining experimental conditions had to be randomised across subjects in order to rule out the effect of the sequence in which the conditions were performed on the obtained results. Similar considerations applied to the sequence of the presentation and to the sizes of the sets to be counted in subsequent trials within each experimental condition. Finally, in order to mitigate the effects of tiredness and fussiness, the psychological experiment had to be split into sessions which took place on different days, as well as the duration of a single session had to be relatively short, what limited the amount of data that has been gathered. All these considerations are completely irrelevant to the design of a simulation experiment with artificial neural networks, because their training is perfectly controlled, and the order of presentation of the test trials does not affect the results in any way.

Although, as has been shown above, the intrinsic properties of the employed artificial neural network framework allow to considerably simplify the experimental protocol (cf. figures 6 and 15), they also require some specific considerations which are not relevant in the case of children. First, since an artificial neural network cannot be expected to perform well on the input patterns it has not been trained on, in the performed simulations the models had to undergo bespoke training in every considered experimental condition (figure 15). As the result, the evaluation across the experimental conditions for a single 'subject' was performed, strictly speaking, on different neural networks. The argument that this nevertheless constitutes the evaluation of the same subject is based on the fact that the training for each condition starts with the exact same weights of the connections, obtained prior to the network extension as the result of the preliminary training stage. Since the experimental

conditions for one subject do not affect each other, as a side benefit it is possible to analyse the results with the use of the repeated-measures approach, which has more discriminative power. Second, in contrast to the study of Alibali and DiRusso (1999) where the arrangement of items in the set to be counted was always the same for the same number (as the result of the constant spacing between the items), the artificial neural network architecture employed herein requires the spatial arrangement of items for a fixed number to be randomised. As explained in section 5.2.1, should this not be the case, the neural network would quickly learn to associate the locations of the items with their number and obtain very high ‘counting’ accuracy scores without actually counting. As the consequence, the evaluation protocol needs to include more test trials, in order to mitigate the effects of increased variability resulting from the spatial randomisation. Finally, the simulations conducted for the needs of the present thesis considered smaller numbers than the reference experimental study (1–10 versus 7–17). The main reason behind this are the considerations connected with the combinatorial properties of the neural network input patterns (the number of different arrangements of items grows exponentially with the largest considered number) and the duration of the training required as the result (cf. section 5.2.3). On the other hand, it can be argued that it would not be correct, from the methodological point of view, to assess the performance of the model on larger numbers without first taking into account what happens within the preceding number range.

There are three more differences between the considered real-life and simulated experimental protocols which are worth pointing out. First, with the artificial neural networks it is possible to simulate scenarios which are not realisable in the psychological experiments with children. More specifically, this applies to the ‘gesture-only’ conditions in the simulation experiments described in sections 6.3 and 6.4, where the counting accuracy of the neural network with only the proprioceptive input available was assessed. These experimental conditions provided vital pieces of information about how the input signals are exploited by the model, which in turn

made it possible to provide answers to the research questions 2 and 3. The rhythmic gestures, considered in the context of the latter research question, can be seen as another example of a situation which does not have an exact equivalent in children, at least in the case of active gestures. The two remaining differences are caused by deliberate design decisions. As discussed in section 6.3.2, since the simulations described in chapter 6 consider only the case of the counting gestures being the input to the model rather than its output, the ‘active gestures’ conditions from the study of Alibali and DiRusso (1999) were not reproduced herein. Similarly, the distinction between touching and pointing-only counting gestures has not been made in my simulations. Although certainly realisable, and in fact attractive from the point of view of embodied robotic modelling, it turned out not to be possible to include the investigation of the effects of the tactile information on the counting accuracy in the scope of the present thesis.

Summarising, while the majority of the differences between the experimental protocols is irrelevant for the obtained results, some of them (such as the different number ranges and not reproduced real-life experimental conditions) already suggest the possible areas for future work. Crucially however, since both experiments use the same task to assess the counting accuracy (HM task without considering the final cardinal response, but only the trajectory of the counting gestures and the three-way correspondence between the items, gestures, and recited number words), since the counting accuracy can be expressed relatively to the number of test trials and is averaged across the considered range of numbers, and since both studies employ identical counting accuracy assessment criteria, the obtained aggregate results can be meaningfully compared.

9.1.2 Discussion of the Results Connected with the Research Question 1

As it has been already pointed out in section 6.1.4, the obtained behaviour of the proposed model with respect to the progress of the learning of the count list did

not match very well with what is observed in children. Namely, the model did not exhibit the stochastic within-subject effects, such as the unstable non-conventional portion of the count list. Also, the model did not have the tendency to emit stable non-conventional sequences — with only few exceptions, at every time step of the simulation the designed neural network produced either a correct number word, or remained silent. This disparity between the behaviour of the model and that of children shows that the employed artificial mechanism of learning of sequences does not capture the properties of human rote learning well. As discussed above in connection with the experimental protocol differences, one of the reasons behind this is the lack of stochasticity in the model's output, once the training is completed. In other words, the employed neural network framework, by its nature, simply does not exhibit within-subject stochasticity. This may suggest that a degree of caution must be taken when generalising the answer to the research question 1 obtained in the course of the present experiments to human cognition. In order for the conclusions from the simulation to be more definitive, the internal structure of the model, and possibly the employed representations, should be revised with a better match to the experimental observations in view.

In addition, it is important to stress that the results which were taken as the answer to the research question 1 hold strictly within the applied training and testing regime. Note, that the crucial element of the experiment was the comparison of the counting accuracy obtained at the end of the neural network training, and the training was interrupted arbitrarily after a fixed amount of training epochs. The duration of the training has been determined based on informal (and therefore subjective) inspection of its progress before the actual experiment. Since no formal stop condition (for instance based on the trend of the training error) has been employed, there is no direct evidence that, given enough training time, the network which learnt to count and to recite number words at the same time could not ultimately achieve the equivalent counting accuracy as the network which has been equipped with the ability to recite the count list beforehand. Therefore, the following altern-

ative methodology would perhaps be more elegant in this situation. A formal stop criterion based on the convergence of the training could be defined and applied in both experimental conditions. Then, once the training for both networks has been completed, one would compare the final counting accuracy, and, in case of a tie, the amount of training it took each network to achieve it. The advantage over the approach used herein would be the obtaining of a quantitative measure of how much faster the network in one experimental condition learns than in the other. The disadvantage would be unpredictable and potentially much longer time necessary to complete the simulations. Crucially however, these considerations do not nullify the results obtained with the use of the strategy that has been adopted in the present thesis, which is sufficient to prove that in one experimental condition the training proceeds faster than in the other, in the qualitative sense.

9.1.3 Discussion of the Results Connected with the Research Questions 2 and 3

As in the case of the research question 1, the results of the simulations which addressed the following two research questions also did not agree with all data obtained with human participants. First, although the proposed model, just as children, achieved consistently higher counting accuracy with gestures than without them, the magnitude of the increase was more prominent than what is observed experimentally (see figure 25 on page 185). Second, and more disturbingly, the performance of the neural network was not statistically affected by the size of the set to be counted, whereas the so-called ‘problem-size effect’ is prevalent not only in counting but also in many other aspects of human numerical cognition (Zbrodoff & Logan, 2005). Third, the distribution of the frequencies of the counting errors committed by the proposed model is not in agreement with that of children, even if only the more appropriate ‘puppet conditions’ are taken into account. As discussed in section 6.3.4, these results are the evidence of some important inherent limitations of the approach employed to model learning to count. Therefore, they also suggest

that the extent to which the conclusions made in the context of research questions 2 and 3 can be extrapolated to human mathematical cognition should be limited.

An important part of the future work with respect to the research questions 2 and 3 is therefore to identify the reasons behind the failures to reproduce the experimental phenomena discussed above and update the model in such a way that the obtained behaviour is more realistic. As it has been hypothesised earlier in this thesis, it is most likely the relative simplicity of the model, its inherently discrete nature with respect to time, and also possibly the employed verbal representation which prevents it from exhibiting all phenomena connected with learning to count in sufficient detail. Especially the results obtained with the counting error patterns suggest that an altogether different choice of the modelling framework may be necessary in order to capture all important aspects of human numerical development. The experimental data indicate that the majority of the counting errors made by children (when they point by themselves) are the skip and double count errors (see figure 27 on page 186). Such errors are likely to be caused primarily by the children's struggle to synchronise the production of words and gestures accurately. Modelling the lack of synchrony is not straightforward within the framework of discrete-time processing (adopted herein), because not much effort on part of the neural network is required for the synchrony to be maintained. As the result, in the proposed model the synchronisation errors turned out to be very unlikely to occur. It seems therefore that, ultimately, the future of the modelling of the contribution of gestures to learning to count lies in the domain of continuous time.

An obvious follow-up analysis of the results obtained in chapter 6, which has not been performed herein, is connected with the exact understanding of the mechanism of the contribution of the counting gestures to the counting accuracy of the proposed model. Although the combined results of the simulations described in sections 6.3 and 6.4 suggest that the proprioceptive signal most likely facilitates somehow the extraction of the information from the visual input, at the time of writing the exact way this happens in the neural network has not yet been reverse engineered.

Understanding how the contribution of gestures works in the model is of course important, because it should be helpful in coming up with testable predictions and in pinning down the nature of the analogous phenomenon observed in children.

Several improvements in the context of the simulations aimed at answering the research questions 2 and 3 could also be made in terms of the way the robotic platform has been employed. In particular, one specific difficulty which often appears in the studies that use robots is the inevitable variability in the input and output signals. Ideally therefore, every robotic model should be demonstrated to be robust with respect to such motor noise. Should that not be the case, this would be a serious blow to any obtained results. In the present simulations, only one set of the proprioceptive data has been gathered with the use of the iCub robot and used to construct the representation of the counting gestures which was fed to every created instance of the proposed model. Thus, as such, the issue of the robustness of the results with respect to noise has not been addressed. This could have been done relatively easily by repeating the procedure described in section 5.2.2 an appropriate number of times and using a different set of proprioceptive data in every repetition of the experiment. This way, the variability in the proprioceptive signal would be an additional contribution to the between-subject variability. The fact that the problem of motor noise could have been addressed so easily, but nevertheless this has not been done, is one of the most serious shortcomings of the present work in the context of the discussed research questions. Fortunately, the magnitude of the observed effects is large enough to allow to predict that additional variability in the proprioceptive signal should not invalidate any of the most important findings; however, whether this is indeed the case will have to be investigated experimentally in future work.

Another issue about the proposed model that might be pointed out in connection with the use of the robotic body is that the counted items are assumed to be located only within a limited number of designated locations, the spatial positions of which are fixed with respect to the robot body, whereas children appear to be able to

enumerate items located anywhere in their operational space. In technical terms, this means the proposed model does not exhibit motor invariance. There is however a number of facts which should be pointed out in this context. First, it is worth to recall that, for a fixed collection size, the model is expected to be able to enumerate the items in *any* possible arrangement in which they can appear in the 20 locations represented by the units of the visual input. Note, that the number of available locations is twice as large as the maximum considered number of items, what allows for adequate ‘room’ even in the extreme case. In fact, as discussed in section 9.1.1, spatial randomisation is required for the model to work correctly. Therefore, the proposed neural network is actually trained to be invariant to the positions of the items, but of course within the limits of the assumed operational space (note that for 5 items there are 15,504 *different* arrangements in which they may appear on 20 spatial locations). This changes the question from ‘whether the model exhibits motor invariance’ to ‘whether the model exhibits motor invariance to a sufficient extent’. Second, I am not aware of conclusive evidence for absolute motor invariance in the case of children in the context of counting in the considered age range. In fact, available evidence, such as the significant effect of the arrangement of the items on children’s counting accuracy (Beckwith & Restle, 1966), argues *against* the hypothesis of absolute motor invariance in this context. Additionally, since the set-ups of the experimental studies usually attempt to rule out any effects of the variability of the arrangement of the items as undesirable (recall that Alibali & DiRusso, 1999, used items of regular and consistent size, carefully located at fixed intervals in a controlled location with respect to the child, and even provided specific instructions as of what children should do with their hands when not gesturing), the lack of ‘absolute’ motor invariance in the model does not pose a big problem from the point of view of the obtained results.

Finally, an important avenue, in which the simulations conducted within the scope of research questions 2 and 3 must be extended, is connected with the relaxation of the assumption about the correctness of the proprioceptive information

provided as an input to the model, and with the delegation of the task of producing the gestures to the model. As evident from the model description in chapter 5, this can be easily incorporated in the proposed neural network architecture, however the time constraints did not allow this to be realised in the present thesis. There are several reasons, some of them discovered only in the course of the conducted simulations, why exploring the behaviour of the model in the ‘gestures as output’ configuration should be pursued in future work. First, as mentioned above when comparing the experimental protocols, such a set-up would correspond more closely to the experiments with children, in which active gestures (performed by the child itself) were investigated; the model configuration considered herein should rather be viewed as corresponding to the ‘passive gestures’ conditions, in which children observe somebody else’s pointing. This is an important point, because allowing the model to produce the counting gestures opens the way for more types of counting errors to appear, what is connected with one of the major weaknesses in the obtained results identified earlier. Neural networks producing gestures as output could be evaluated using a very similar experimental methodology as the one applied in the simulations described in chapter 6. An additional possibility the ‘gestures as output’ set-up would open, is the evaluation of the counting behaviour of a model, which has been trained in the presence of the counting gestures, after the gestures are taken away, since removing an output from a neural network affects it much less than removing some of its inputs. This way, an attempt to realistically model the transfer of the motor competence to a conceptual one, which is evident in children in that after their counting competence is mature enough they do not need gestures any more (see section 2.3.2), could be made.

9.2 Future Work — Research Question 4

Although obtained with a different neuro-robotic model to the one which has been used to answer the research questions 1–3, a number of similar problems can be

pointed out with the results which led to the final conclusions presented in connection with the research question 4. As the review of the literature on computational modelling in mathematical cognition in chapter 3 reveals, the model of the acquisition of the spatial-numerical associations introduced in chapter 7 is not the first one to focus on this phenomenon. In fact, it can be considered an embodied extension of the antecedently published works of Gevers, Verguts et al. (2006) and Chen and Verguts (2010). As discussed in section 8.1.4, one of my main results is the demonstration that the assumptions of Chen and Verguts with respect to the sources of the SNARC effect in their model are plausible, what can be regarded as validation of their approach. It is however worth to point out, that the publication of Chen and Verguts impresses with the amount of behavioural data which the model described therein reproduced, and many of these scenarios have not been simulated here. Therefore, in order for the conclusion about the validation of the referenced model to be fully justifiable, it needs to be demonstrated that the development process proposed in chapter 7 leads to the reproduction of not only the fundamental ones, but all phenomena that the original work was able to account for. In fact, considering the embodied character of some of the tasks simulated by Chen and Verguts, such as the effects of the spatial neglect or the physical line bisection task, it would be quite an attractive endeavour from the point of view of neuro-robotic modelling.

Before turning to new experiments, it would be however appropriate to address the issues which can be identified with the results obtained so far. First such an issue is the fact, that, in contrast to the referenced works, the assessment of the patterns of the response times obtained in the simulations described in chapter 8 was based solely on visual inspection and qualitative comparison with human data, rather than formal quantitative analysis. More specifically, the obtained model chronometry has not been regressed against that of humans (as has been done by Chen & Verguts, 2010). Although this could require considerable effort, as quite likely this would involve adjusting multiple parameters of the model before a good

quality fit is found (Grossberg & Repin, 2003), such an analysis could provide useful insight into the similarity of the behaviour of the model to that of humans, what in turn should highlight the ways in which the employed approach could be improved. Of course the imperfections in the results which have been identified already, such as the border effect visible in figure 45a, need to be addressed as well.

Unfortunately, one of the biggest problems noted in connection with the first group of the simulation experiments in the context of the robotic modelling methodology, also applies to the experiments conducted in connection with the spatial-numerical associations. That is, also here only one set of the motor data has been gathered (in the process of simulating the motor babbling) and re-used multiple times in subsequent repetitions of the experiment. Therefore, also in this case, the obtained results have not been demonstrated to be robust with respect to motor noise. In addition, unlike in the former experiment, in the part III of this thesis the simulated version of the robot has been employed instead of the real one. An attempt to reproduce the experiments using the real robot was undertaken, but became delayed due to the problems connected with the differences between the real robot and its simulated equivalent. As the result, this has not been completed within the time frame of the present thesis and remains as another area of future work.

Finally, a factor which needs to be kept in mind in connection to the answer provided to the research question 4 is that there are ways, in which the proposed model of spatial-numerical associations could be improved in terms of the corpus of the properties of the considered phenomena it is able to account for. This is important, because the less realistic the behaviour of the model is, the more difficult it is to generalise from the conclusions of the simulation experiments. One prominent characteristic of the SNARC effect which has not been captured by any of the past models nor by the one proposed here, is its notable plasticity. As reviewed in chapter 2, section 2.2.3, the following aspects of the SNARC effect's flexibility can be distinguished:

- with respect to the considered number range (the same number perceived as small or large depending on the context);
- orientation of the mapping (left-to-right, right-to-left, bottom-to-top, etc.);
- topology of the mapping (directional versus ‘clock face’);
- presence for stimuli other than numbers;
- connection with response side (and not response hand);
- short- and long-term shaping of the effect (influence of culture versus task instructions).

The multi-dimensional plasticity of the SNARC effect is so striking that the results of any efforts to model it which do not take this flexibility into account, immediately run into risk of being questioned. All existing models fail, in general, to reproduce any of the above aspects of the SNARC effect plasticity, if one requires it must be accounted for within a single instantiation of a model. Some of them, like the flexibility with respect to the number range or with respect to the orientation and topology of the number-space mapping, can in principle be obtained using the model proposed herein. The former would however require changes in the structure of the neural network, such as the adjustment of the sizes of the input and semantic layers in the What pathway. The latter in turn, would require the training process to be repeated with differently constructed training data sets. Even then, the model would be constrained to only one number-space mapping at a time, therefore the short-term plasticity of the SNARC effect would not be exhibited.

In order to at least partially account for the plasticity of the SNARC effect, the model proposed in chapter 7 can be extended to be able to encompass multiple number-space associations and to evoke them in short-term by the context of the task. It is also possible to make the SNARC effect be bound to the response side rather than to the response hand. The concept of such a model has been formulated as part of the work for this thesis, however the time constraints did not permit it

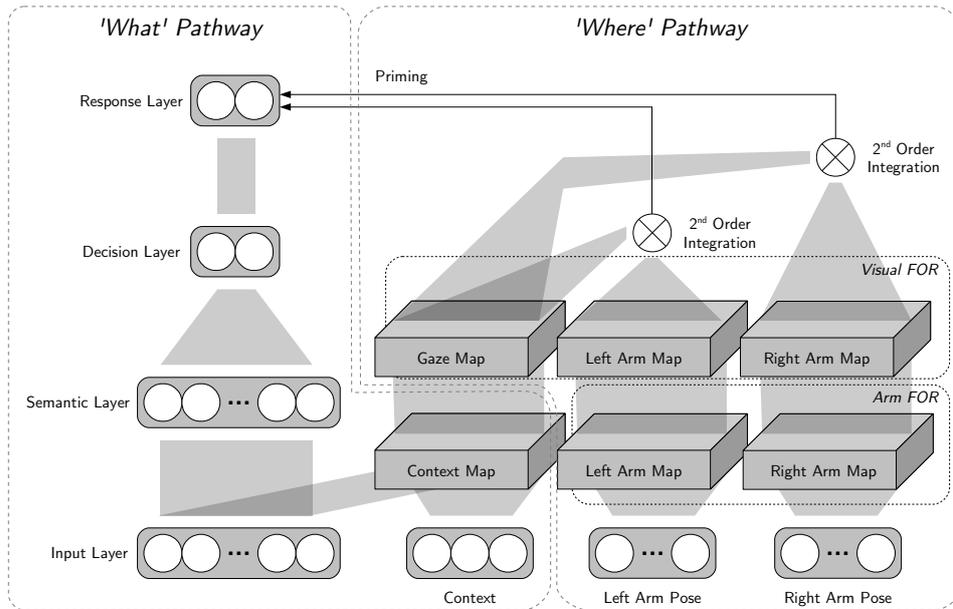


Figure 48: Concept of the extension of the model of the acquisition of spatial-numerical associations.

to be implemented and simulated. The schematic diagram of the extended model is shown in figure 48. The neural network features an extended Where pathway, which at all times represents the current poses of the robot's arms. These are translated into representations which use the same frame of reference as the visual map. The form of the association of numbers with space in the model is dependent on the task context, what is implemented by introducing an intermediate self-organising map between the input layer and the gaze map (alike to the Epigenetic Robotics Architecture of Morse et al., 2010). This intermediate map routes the propagation of neural activations between the input and gaze maps, effectively allowing multiple 'number forms' to be stored in one model and to switch between those instantly by changing the values of the context inputs. The amount of priming of the left- and right-handed response is determined through the integration of the information in the gaze and arm maps (what would most likely require second-order processing). As the result of this, changing the locations of the arms of the robot (e.g. to achieve crossed-hands conditions) should change the priming of the responses accordingly.

Theoretically, the model shown in figure 48 should be able to reproduce several aspects of the SNARC effect not accounted for by any of the earlier models. However,

since the implementation and simulations of this model fell out of scope of the thesis, confirming whether this is indeed the case is another avenue for future work.

9.3 Future Prospects of Developmental Cognitive Robotics in Mathematical Cognition

The central theme throughout this thesis, and one of its major contributions in the general sense, was the first large-scale application of the principles of the developmental cognitive robotics methodology in the context of the study of human numerical capabilities. In this concluding section I am going to adopt a slightly wider perspective, and briefly review some topics in mathematical cognition not necessarily connected with learning to count, which appear promising from the point of view of the investigation with the use of embodied robotic modelling, and thus have significant potential for becoming the topics of such research in the future.

The first topic worth considering is connected with the question why modelling counting in robots is important — not as much in terms of the benefits it provides for cognitive modelling, what has been highlighted quite frequently in the present work — but, *from the point of view of robotics*. The answer to this question is not trivial. Notably, there is a stark contrast between how counting is performed by humans and how a roboticist with little interest in the cognitive aspects of the matter would most likely attempt to solve an analogous task. Counting in humans is, in general, a *sequential* process. In turn, the most common and ‘natural’ way of the analysis of digital images in the context of the state of the art of computer (robotic) vision is based on the processing of information that are available *all at once*, in a frame buffer that covers a relatively wide-angle view of the scene. From the point of view of computer vision, the main difficulty to overcome in order to enumerate items in a scene is to correctly identify the areas of the two-dimensional image that are occupied by the objects to be counted. Assuming this can be done, the numerical information is already implicitly available in the form of the number of the segmented

regions. Considering this, it is quite likely that a computer vision engineer would not attribute much importance to the modelling of *sequential* counting in a robot — at present, it simply does not seem to be necessary or useful. In the light of this fact, it is worth to dwell for a moment on the possible reasons why humans *do not* extract exact numerical information for large sets instantly, as currently available robots could do? Clearly, it is quite likely that the approach humans take to enumerate items is connected with the way they collect visual information. In fact, as mentioned earlier in this thesis, counting is sometimes explicitly defined in terms of the ‘shifts of attention’ (Beckwith & Restle, 1966; Andres et al., 2007). It is well established that human vision works quite differently from the present-day digital cameras — it is precise in a relatively narrow *foveal* region, and significantly less so in the wider-angle *peripheral* region (Schaeffel, 2007). It is therefore quite plausible to hypothesise that this property of human vision is responsible for the need for performing sequential enumeration in order to establish the exact number of items even in a static scene. However, is the necessity to count merely a *drawback* of having a foveal, rather than ‘global’ vision? Or could it have some advantages in the context of extracting numerical information in the visual modality? Does having a foveal vision (or, more generally, selective attention) *facilitate* the acquisition of the abstract concept of number, for instance by forcing us to enumerate static visual stimuli in the same way we do with temporal stimuli? Could foveal vision or selective attention even be *necessary* to arrive at the concept of number? The importance of foveal vision in the context of extraction of numerical information appears to be a promising avenue for future research, due to its strong links with embodiment. Sufficiently precise modelling of the human visual system would of course be an important aspect of such studies.

The second theme in mathematical cognition, in the investigation of which an artificial robotic body could be particularly useful, is finger counting. As mentioned in chapter 2, section 2.2.2, the evidence for the involvement of fingers in numerical thinking is abundant (see Fischer & Brugger, 2011, for review). Fayol and Seron

(2005, pp. 15–16) put forward the following reasons why finger counting is a powerful candidate for supporting the grounding of the symbolic understanding of numbers in the pre-verbal quantitative knowledge:

- a particular configuration of raised fingers constitutes a discrete, symbolic representation of a numerosity;
- similarly to language, fingers can be considered an abstract representation, as the same configuration of fingers can represent a specific number of items of any kind;
- in contrast to language, the finger representation is iconic, that is the raised fingers can be put in one-one correspondence with the set they represent;
- the fingers exceed in magnitude the system of exact representation of small numbers connected with our object individuation abilities (cf. section 2.1.1);
- raising and lowering fingers is directly related to adding and taking away items to and from a set, potentially supporting the understanding of addition and subtraction;
- a configuration of fingers can be used as a memory aid;
- joint use of fingers of two hands may be helpful in understanding the concept of numeral systems with base.

It is quite conceivable to imagine experiments in which an attempt would be made to ground symbolic numerical knowledge of a robot in the motor competences connected with the robot's hands and fingers. Would it be helpful to link numerical representations with the motor representations of hand postures? Would it be possible to transfer the relations that exist between hand configurations (such as inclusion) to the numerical domain? Would it be possible to teach the robot simple arithmetic based on finger movements? There is a vast amount of phenomena connected with finger counting that could be explored. In fact, at the time of writing, first

neuro-robotic modelling experiments in this context begin to appear (De La Cruz, Di Nuovo, Di Nuovo & Cangelosi, to appear). For such research, robots equipped with end effectors that highly resemble human hands (such as the iCub robot used in the present study) are particularly suitable.

The last topic in mathematical cognition I would like to put forward as having, in my opinion, a high potential of becoming the subject of cognitive developmental robotics experiments are the hypotheses about the embodied sources of various mathematical concepts formulated by Lakoff and Núñez (2000, see section 2.2.1). Among the four grounding metaphors for arithmetic proposed by Lakoff and Núñez, one appears to be especially compelling and relevant from the point of view robotics — namely, the *Arithmetic Is Motion Along a Path* metaphor. According to this metaphor, concepts such as zero, numbers, greater and less than, as well as elementary arithmetic operations can be grounded in such primitives related to motor control as the origin of the motion, the locations along the motion path, distance to the origin along the path and moving away from and toward the origin. This metaphor is particularly interesting in the considered context because it is hard to think of any issue more fundamental and central to robotics than motion and motor control. Moreover, as reviewed in section 4.2 of this thesis, motor control, in addition to being vitally important to ‘conventional’ robotics, has also been investigated in a large body of studies in cognitive developmental robotics, from the very beginning of the existence of this discipline (Lungarella et al., 2003). The endeavour to study embodied sources of arithmetic with the aid of robotic modelling could focus on the following intriguing questions. First of all, how conceptual metaphor could be ‘implemented’ in a robot control system — be it symbolic or sub-symbolic? Then, would it be possible to ground a (possibly simplified) version of ‘robotic arithmetic’ in the motor primitives of the robot’s control system? Would this make it possible to teach the robot the numbers or the zero concept? Would such a robot be able to learn to add and subtract based on the analogy to the motions in its repertoire? Would it be possible to combine the conceptual knowledge from multiple source

domains, for instance including already mentioned finger counting? As a matter of fact, it is possible to call upon already existing studies in cognitive robotics that have looked at some related issues. For instance, Cangelosi and Riga (2006) investigated how language (more specifically, words such as ‘grab’, ‘push’, ‘pull’ or ‘carry’) can be grounded in the motor actions of a simulated robot. More recently, Stramandinoli et al. (2012) looked at a similar topic with the use of the iCub robot, in the context of more abstract concepts, such as ‘accept’, ‘reject’, or ‘keep’. Analogous research could be performed to investigate the grounding of mathematical concepts with the use of metaphorical mechanisms. Finally, studies exist that have tackled the conceptual blending and the 4Gs of Lakoff and Núñez from the computational perspective (Guhe et al., 2011). Bringing the results of these efforts together and investigating further the metaphorical grounding of mathematical concepts holds an exciting promise of being able to build a robot not only capable of learning arithmetic but also one that could be argued to *understand* it.

Of course, the three themes discussed above do not exhaust the potential further uses of cognitive robotics in the investigation of mathematical cognition. The importance of the embodied approach to cognitive modelling, not only in the context of numerical knowledge but for cognitive science in general, may be expected to increase along with the complexity of the phenomena being investigated, and with the degree to which the motor representations and actions are involved. Since the amount of evidence for the embodied character of various cognitive processes is constantly growing, I expect the importance of robotic modelling in this context to continue to grow as well. It seems that the application of robotic models in cognitive science has a bright future and will contribute to our understanding of the phenomena that involve both our brain and body.

Appendix A

Equations Describing the Model of the Acquisition of Spatial-Numerical Associations

The model described in chapter 7 of this thesis is implemented in the firing rate framework in which the dynamics of the activity of the units of the neural network are described by a system of ordinary differential equations (cf. chapter 4 section 4.4.2). In the present appendix the equations of the model in all used configurations are given, in order to supplement the verbal description. The framework of the implementation of the discussed model is analogous to that used by Verguts et al. (2005), Gevers, Verguts et al. (2006), and Chen and Verguts (2010).

A.1 Notation

In the subsequent sections of the present appendix the following notation is assumed:

$x_i^L(t)$ activation value of the i -th unit of the layer L at time t ;

$W^{L,M}$ matrix of connection weights from the layer L to the layer M ;

$w_{i,j}^{L,M}$ weight of the connection from the unit j in layer L to the unit i in layer M
(an element of the connection weights matrix $W^{L,M}$);

a_i i -th element of the vector \mathbf{a} ;

$\mathbf{0}_n$ n -dimensional null vector.

Time is represented in arbitrary time units of the numerical integrator used to solve the system of equations. The initial conditions are such that $x_i^L(0) = 0$ for all L and i .

A.2 Number Comparison Task

In simulations of the number comparison task (i.e. in simulations of the number size and numerical distance effects in section 8.2 and of the SNARC effect in section 8.3), the neural network consists of the following layers of units (cf. figure 42):

- two input layers (designated INP_1 and INP_2 , 15 units each);
- two semantic layers (SEM_1 and SEM_2 , 15 units each);
- decision layer (DEC , 2 units);
- response layer (RES , 2 units);
- gaze map (GAZ , 49 units);
- left arm map (LFT , 49 units);
- right arm map (RGT , 49 units).

The inputs to the model (two numbers being compared) are represented using two 15-dimensional column vectors \mathbf{a} and \mathbf{b} , using one-hot coding. The activity of the units of the input layers is given by:

$$\frac{d}{dt}x_i^{INP_1}(t) = -x_i^{INP_1}(t) + a_i \quad \text{for } 1 \leq i \leq 15 \quad (\text{A.1})$$

$$\frac{d}{dt}x_i^{INP_2}(t) = -x_i^{INP_2}(t) + b_i \quad \text{for } 1 \leq i \leq 15 \quad (\text{A.2})$$

The activity of the units of the semantic layers is given by:

$$\frac{d}{dt}x_i^{SEM_1}(t) = -x_i^{SEM_1}(t) + \sum_{j=1}^{15} w_{i,j}^{INP,SEM} x_j^{INP_1}(t) \quad \text{for } 1 \leq i \leq 15 \quad (\text{A.3})$$

$$\frac{d}{dt}x_i^{SEM_2}(t) = -x_i^{SEM_2}(t) + \sum_{j=1}^{15} w_{i,j}^{INP,SEM} x_j^{INP_2}(t) \quad \text{for } 1 \leq i \leq 15 \quad (\text{A.4})$$

where $W^{INP,SEM}$ is the matrix of weights of connections between an input layer and a semantic layer. The desired semantic encoding of numbers described in section 7.2.1 of chapter 7 is obtained by assuming:

$$w_{i,j}^{INP,SEM} = e^{-|j-i|} \quad \text{for } 1 \leq i \leq 15, 1 \leq j \leq 15 \quad (\text{A.5})$$

The activity of the units of the decision layer is given by:

$$\begin{aligned} \frac{d}{dt}x_i^{DEC}(t) = \\ -x_i^{DEC}(t) + \sum_{j=1}^{15} w_{i,j}^{SEM_1,DEC} x_j^{SEM_1}(t) + \sum_{k=1}^{15} w_{i,k}^{SEM_2,DEC} x_k^{SEM_2}(t) \quad \text{for } 1 \leq i \leq 2 \end{aligned} \quad (\text{A.6})$$

where $W^{SEM_1,DEC}$ and $W^{SEM_2,DEC}$ are the matrices of weights of connections between the semantic layers and the decision layer. The values of these weights are obtained throughout the supervised training process described in section 7.3.4 (see also figures 41b and 41c).

The activity of the units of the gaze map is given by:

$$\begin{aligned} \frac{d}{dt}x_i^{GAZ}(t) = \\ -x_i^{GAZ}(t) + \sum_{j=1}^{15} w_{i,j}^{INP,GAZ} x_j^{INP_1}(t) + \sum_{k=1}^{15} w_{i,k}^{INP,GAZ} x_k^{INP_2}(t) \quad \text{for } 1 \leq i \leq 49 \end{aligned} \quad (\text{A.7})$$

where $W^{INP,GAZ}$ is the matrix of weights of connections between an input layer and the gaze layer. These connections represent the actual spatial-numerical association (i.e. the association of small numbers with the left side of space and of large numbers with the right side of space) and are obtained through the process of Hebbian learning described in section 7.3.3 (see also figure 40).

The activity of the units of the left and right arm maps is given by:

$$\frac{d}{dt}x_i^{LFT}(t) = -x_i^{LFT}(t) + \sum_{j=1}^{49} w_{i,j}^{GAZ,LFT} x_j^{GAZ}(t) \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.8})$$

$$\frac{d}{dt}x_i^{RGT}(t) = -x_i^{RGT}(t) + \sum_{j=1}^{49} w_{i,j}^{GAZ,RGT} x_j^{GAZ}(t) \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.9})$$

where $W^{GAZ,LFT}$ and $W^{GAZ,RGT}$ are the matrices of weights of connections between the gaze layer and the left arm map, and between the gaze layer and the right arm map, respectively. These weights implement transformations between the frames of reference associated with the spatial maps and are established in the first stage of the model development process, described in section 7.3.1 (cf. figure 39).

Finally, the activity of the units of the response layer is given by:

$$\begin{aligned} \frac{d}{dt}x_i^{RES}(t) = & -x_i^{RES}(t) + \sum_{j=1}^2 w_{i,j}^{DEC,RES} x_j^{DEC}(t) \\ & + \sum_{k=1}^{49} w_{i,k}^{LFT,RES} x_k^{LFT}(t) + \sum_{l=1}^{49} w_{i,l}^{RGT,RES} x_l^{RGT}(t) \quad \text{for } 1 \leq i \leq 2 \quad (\text{A.10}) \end{aligned}$$

Here, $W^{DEC,RES}$ is assumed to be equal either to the matrix W_{reg} given in equation 7.3 or to W_{inv} (equation 7.4), what implements two reverse response mappings (i.e. give left-handed response when the first number is larger vs. first number is smaller), required in the simulations of the SNARC effect. Weights $W^{LFT,RES}$ and $W^{RGT,RES}$ are set-up as:

$$w_{1,i}^{LFT,RES} = w_{2,i}^{RGT,RES} = \alpha \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.11})$$

$$w_{2,i}^{LFT,RES} = w_{1,i}^{RGT,RES} = -\alpha \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.12})$$

where α is a parameter of the model which determines the strength of the priming of the response by the Where pathway. As noted in their respective descriptions, the simulations of the number comparison task reported in the present thesis assumed $\alpha = 25$.

A.3 Parity Task

In the simulations of the SNARC effect in the number parity task (section 8.3), the neural network consists of the following layers of units (cf. figure 44):

- input layer (*INP*, 15 units);
- semantic layer (*SEM*, 15 units);
- decision layer (*DEC*, 2 units);
- response layer (*RES*, 2 units);
- gaze map (*GAZ*, 49 units);
- left arm map (*LFT*, 49 units);
- right arm map (*RGT*, 49 units).

The only difference between the configuration of the model used in the parity task and the one used in the number comparison task is the lack of the duplicated input and semantic layers in the former. This results from the fact that, for parity judgement, only one number is required as input. The dynamics of the model in the parity task are therefore identical to the number comparison task, except for the following. The only input to the model is the vector \mathbf{a} , representing the input number using one-hot coding. Activity of the layers *INP* and *SEM* are as those given by equations A.1 and A.3, respectively. Since the layers *INP*₂ and *SEM*₂ are not present, equations A.2 and A.4 are left out. Equation A.6 takes the form:

$$\frac{d}{dt}x_i^{DEC}(t) = -x_i^{DEC}(t) + \sum_{j=1}^{15} w_{i,j}^{SEM,DEC} x_j^{SEM}(t) \quad \text{for } 1 \leq i \leq 2 \quad (\text{A.13})$$

where the connection weights matrix $W^{SEM,DEC}$ is obtained using the supervised training process described in section 7.3.4 (see figure 41a). The activity of the remaining layers of units follows the respective equations from section A.2 (i.e. equations A.7 through to A.12). However, as mentioned in section 8.3, in the parity task simulations it has been assumed that $\alpha = 50$.

A.4 Visual Target Detection Task

In the simulations of the Posner-SNARC effect which involve the visual target detection task (see section 8.4, and figure 46) the neural network consists of the following layers:

- input layer (INP , 15 units);
- gaze map (GAZ , 49 units);
- response layer (RES , 2 units).

The most prominent differences in comparison to the model configurations described above are that in this case inputs to the neural network change over time, and that in addition to the symbolic numerical input, visual information representing the target to be detected is also provided.

Let \mathbf{a} be a 15-dimensional vector, encoding the number being used as the priming stimulus using one-hot coding. The activity of the units in the input layer is given by:

$$\frac{d}{dt}x_i^{INP}(t) = -x_i^{INP}(t) + a_i(t) \quad \text{for } 1 \leq i \leq 15 \quad (\text{A.14})$$

where $a(t)$ is a 15-dimensional function defined as:

$$a(t) = \begin{cases} \mathbf{a} & \text{for } t < t_{pri} \\ \mathbf{0}_{15} & \text{otherwise} \end{cases} \quad (\text{A.15})$$

where t_{pri} is the duration of the priming stimulus expressed in the arbitrary time units of the numerical integrator.

The activity of the units in the gaze map is given by:

$$\frac{d}{dt}x_i^{GAZ}(t) = -x_i^{GAZ}(t) + \sum_{j=1}^{15} w_{i,j}^{INP,GAZ} x_j^{INP}(t) + v_i(t) \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.16})$$

where $v(t)$ is a 49-dimensional function which represents the response of the gaze spatial map to the appearance of the visual target to be detected. The function $v(t)$ is defined in terms of two 49-dimensional column vectors \mathbf{v}_l and \mathbf{v}_r which contain the activation values of the units of a fully-developed gaze SOM computed using the equation 7.2 in response to the input vectors that represent the left-sided and the right-sided visual target (designated as \mathbf{x}_l and \mathbf{x}_r), respectively. \mathbf{x}_l and \mathbf{x}_r are defined as points in the gaze SOM input space with vertical coordinates equal to the middle of the vertical span of the gaze SOM training data, and the horizontal coordinates equal to the left- and right-most location in the gaze SOM training data, respectively. $v(t)$ can therefore be written down as:

$$v(t) = \begin{cases} \mathbf{0}_{49} & \text{for } t < t_{pri} + t_{delay} \\ \mathbf{v}_l & \text{for } t \geq t_{pri} + t_{delay} \text{ and left-sided visual target} \\ \mathbf{v}_r & \text{for } t \geq t_{pri} + t_{delay} \text{ and right-sided visual target} \end{cases} \quad (\text{A.17})$$

where t_{delay} is the delay between the priming stimulus and the appearance of the visual target. As reported in chapter 8, section 8.4, in the simulations described therein t_{pri} was assumed to be equal to 2 and the model was tested with t_{delay} equal to 0.2, 1.0, 1.8, and 2.6.

The activity of the response units is given by:

$$\frac{d}{dt}x_i^{RES}(t) = -x_i^{RES}(t) + \sum_{j=1}^{49} w_{i,j}^{GAZ,RES} x_j^{GAZ}(t) \quad \text{for } 1 \leq i \leq 2 \quad (\text{A.18})$$

where $W^{GAZ,RES}$ is a matrix of connection weights which implements the visual target detection. This matrix is constructed as follows. Let W^D be such a 1-column matrix, that:

$$w_i^D = \begin{cases} 1 & \text{if } d_{out}(i, \text{BMU}(\mathbf{x}_l)) < d_{out}(i, \text{BMU}(\mathbf{x}_r)) \\ -1 & \text{if } d_{out}(i, \text{BMU}(\mathbf{x}_l)) > d_{out}(i, \text{BMU}(\mathbf{x}_r)) \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \leq i \leq 49 \quad (\text{A.19})$$

where $d_{out}(i, j)$ is the distance between the SOM units i and j in the SOM output space (i.e. in the SOM topology). In essence, W^D divides the gaze SOM into the left and right ‘hemifields’ by containing the value of 1 for those SOM units which are located closer in the SOM topology to the BMU of the vector \mathbf{x}_l than to the BMU of the vector \mathbf{x}_r (the left hemifield), and -1 for the units for which the opposite is true (the right hemifield). $W^{GAZ,RES}$ is then defined as:

$$w_{i,j}^{GAZ,RES} = \begin{cases} w_j^D & \text{for } i = 1 \\ -w_j^D & \text{for } i = 2 \end{cases} \quad \text{for } 1 \leq j \leq 49 \quad (\text{A.20})$$

The connectivity pattern implemented by the matrix $W^{GAZ,RES}$ means that the left response unit is excited by the left side of the visual field and inhibited by the right side of the visual field, and that the opposite is true for the right response unit.

References

- Abidi, S. S. R. & Ahmad, K. (1997). Conglomerate neural network architectures: The way ahead for simulating early language development. *Journal of Information Science and Engineering*, *13*(2), 235–266.
- Ahmad, K. & Bale, T. A. (2001). Simulation of quantification abilities using a modular neural network approach. *Neural Computing & Applications*, *10*(1), 77–88.
- Ahmad, K., Casey, M. & Bale, T. (2002). Connectionist simulation of quantification skills. *Connection Science*, *14*(3), 165–201.
- Alibali, M. W. & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*(1), 37–56.
- Amit, D. J. (1988). Neural networks counting chimes. *Proceedings of the National Academy of Sciences*, *85*(7), 2141–2145.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.
- Andres, M., Seron, X. & Olivier, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience*, *19*(4), 563–576.
- Antell, S. & Keating, D. (1983). Perception of numerical invariance in neonates. *Child Development*(54), 695–701.
- Arbib, M. A. (2002a). *The handbook of brain theory and neural networks*. MIT Press.
- Arbib, M. A. (2002b). The mirror system, imitation, and the evolution of language. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts*. Cambridge, MA: The MIT Press.
- Asada, M., MacDorman, K., Ishiguro, H. & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, *37*(2–3), 185–193.
- Ashcraft, M. (1987). Children’s knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. Brainerd & R. Kail (Eds.), *Formal methods in developmental psychology: progress in cognitive development research*. New York, NY: Springer Verlag.
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1–2), 75–106.
- Bachot, J., Gevers, W., Fias, W. & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology Science*, *47*(1), 172–183.
- Bächtold, D., Baumüller, M. & Brugger, P. (1998). Stimulus-response compatibility

- in representational space. *Neuropsychologia*, 36(8), 731–735.
- Banks, W. P., Fujii, M. & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3), 435–447.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(04), 577–660.
- Beckwith, M. & Restle, F. (1966). Process of enumeration. *Psychological Review*, 73(5), 437–444.
- Behne, T., Liskowski, U., Carpenter, M. & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, 30(3), 359–375.
- Berch, D. B., Foley, E. J., Hill, R. J. & Ryan, P. M. (1999). Extracting parity and magnitude from arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, 74(4), 286–308.
- Bijeljac-Babic, R., Bertoncini, J. & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, 29(4), 711–721.
- Bisanz, J., Sherman, J. L., Rasmussen, C. & Ho, E. (2005). Development of arithmetic skills and knowledge in preschool children. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 143–162). New York and Hove: Psychology Press.
- Boden, M., Wiles, J., Tonkes, B. & Blair, A. (1999). Learning to predict a context-free language: analysis of dynamics in recurrent hidden units. In *Artificial neural networks, 1999. ICANN 99. ninth international conference on (conf. publ. no. 470)* (Vol. 1, pp. 359–364).
- Brannon, E. M. (2005). What animals know about numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). New York and Hove: Psychology Press.
- Brannon, E. M., Wusthoff, C. J., Gallistel, C. R. & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, 12(3), 238–243.
- Breazeal, C. L. (2002). *Designing sociable robots*. Cambridge, MA: The MIT Press.
- Breckinridge Church, R. & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1), 43–71.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3), 139–159.
- Brybaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, 124(4), 434–452.
- Butcher, J. C. (2008). *Numerical methods for ordinary differential equations* (2nd ed.). Chichester, England: John Wiley & Sons Ltd.
- Butterworth, B. (2000). *The mathematical brain*. London: Macmillan.
- Butterworth, B., Zorzi, M., Girelli, L. & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 54A(4), 1005–1029.

- Caligiore, D., Borghi, A. M., Parisi, D. & Baldassarre, G. (2010). TRoPICALS: A computational embodied neuroscience model of compatibility effects. *Psychological Review*, *117*(4), 1188–1228.
- Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*(1), 1–44.
- Campbell, J. I. D. (Ed.). (2005). *Handbook of mathematical cognition*. Psychology Press.
- Campbell, J. I. D. & Clark, J. M. (1992). *Cognitive number processing: An encoding-complex perspective*. Oxford, England: North-Holland.
- Cangelosi, A. & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive science*, *30*(4), 673–689.
- Cangelosi, A. & Schlesinger, M. (to appear). *Developmental robotics: From babies to robots*. Cambridge, MA: MIT Press.
- Carlson, R. A., Avraamides, M. N., Cary, M. & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology-Learning Memory and Cognition*, *33*(4), 747–756.
- ChaLearn gesture challenge website*. (2013). Retrieved 6 June 2013, from <http://gesture.chalearn.org/dissemination>
- Chen, Q. & Verguts, T. (2010). Beyond the mental number line: A neural network model of number-space interactions. *Cognitive Psychology*, *60*(3), 218–240.
- Church, R. M. & Broadbent, H. A. (1990). Alternative representations of time, number, and rate. *Cognition*, *37*(1–2), 55–81.
- Church, R. M. & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblatt, T. G. Bever & H. S. Terrace (Eds.), *Animal cognition* (pp. 445–464). Hillsdale: Erlbaum.
- Clark, A. (1998). Embodied, situated, and distributed cognition. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 506–517). Malden, MA: Blackwell Publishers.
- Clark, J. M. & Campbell, J. I. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, *17*(2), 204–239.
- Clearfield, M. W. & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, *10*(5), 408–411.
- Clearfield, M. W. & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, *2*(3), 243–260.
- CoDyCo project website*. (2013). Retrieved 3 June 2013, from <http://www.codyco.eu/>
- Colombo, J., Brez, C. C. & Curtindale, L. M. (2013). Infant perception and cognition. In I. B. Weiner, R. M. Lerner, M. A. Easterbrooks & J. Mistry (Eds.), *Handbook of psychology: Developmental psychology* (2nd ed., Vol. 6, pp. 61–89). Hoboken, NJ, US: John Wiley & Sons Inc.
- Cordes, S. & Gelman, R. (2005). The young numerical mind: When does it count? In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 127–142). New York and Hove: Psychology Press.
- Cordes, S., Gelman, R., Gallistel, C. R. & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*(4), 698–707.

- Coventry, K. R., Cangelosi, A., Newstead, S., Bacon, A. & Rajapakse, R. (2005). Grounding natural language quantifiers in visual attention. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *XXVII annual conference of the cognitive science society* (pp. 506–511). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, *13*(1), 21–27.
- Cowan, R., Dowker, A., Christakis, A. & Bailey, S. (1996). Even more precisely assessing children’s understanding of the order-irrelevance principle. *Journal of Experimental Child Psychology*, *62*(1), 84–101.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42.
- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.
- Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, *12*(3), 244–246.
- Dehaene, S. (2002). Single-neuron arithmetic. *Science*, *297*(5587), 1652–1653.
- Dehaene, S. (2003). The neural basis of the weber-fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147.
- Dehaene, S. & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 314–326.
- Dehaene, S., Bossini, S. & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.
- Dehaene, S. & Brannon, E. M. (Eds.). (2011). *Space, time and number in the brain*. San Diego: Academic Press.
- Dehaene, S. & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *J. Cognitive Neuroscience*, *5*(4), 390–407.
- Dehaene, S. & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*(1), 83–120.
- Dehaene, S., Dehaene-Lambertz, G. & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.
- Dehaene, S., Dupoux, E. & Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 626–641.
- Dehaene, S. & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29.
- de Hevia, M. D. & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, *21*(5), 653–660.
- De La Cruz, V., Di Nuovo, A., Di Nuovo, S. & Cangelosi, A. (to appear). *Making fingers and words count in a cognitive robot*.
- Doricchi, F., Guariglia, P., Gasparini, M. & Tomaiuolo, F. (2005). Dissociation between physical and mental number line bisection in right hemisphere brain damage. *Nature Neuroscience*, *8*(12), 1663–1665.
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190.
- Duncan, E. M. & McFarland, C. E. (1980). Isolating the effects of symbolic dis-

- tance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, 8(6), 612–622.
- Eljaik, J., Li, Z., Randazzo, M., Parmiggiani, A., Metta, G., Tsagarakis, N. & Nori, F. (2013). Quantitative evaluation of standing stabilization using stiff and compliant actuators. In *Proceedings of Robotics: Science and Systems*. Berlin, Germany.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Fanello, S. R., Gori, I., Metta, G. & Odone, F. (2013). One-shot learning for real-time action recognition. In J. a. M. Sanches, L. Micó & J. S. Cardoso (Eds.), *Pattern recognition and image analysis* (Vol. 7887, pp. 31–40). Springer.
- Fayol, M. & Seron, X. (2005). About numerical representations: Insights from neuropsychological, experimental, and developmental studies. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 3–22). New York and Hove: Psychology Press.
- Feigenson, L., Dehaene, S. & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314.
- Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research/Psychologische Forschung*, 65(4), 250–259.
- Fias, W., Brysbaert, M., Geypens, F. & d’Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, 2, 95–110.
- Fias, W. & Fischer, M. H. (2005). Spatial representation of numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 43–54). New York and Hove: Psychology Press.
- Fias, W., van Dijck, J.-P. & Gevers, W. (2011). How is number associated with space? the role of working memory. In S. Dehaene & E. M. Brannon (Eds.), *Space, time and number in the brain* (pp. 133–148). San Diego: Academic Press.
- Fischer, M. H. (2003). Spatial representations in number processing — evidence from a pointing task. *Visual Cognition*, 10(4), 493–508.
- Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex*, 44(4), 386–392.
- Fischer, M. H. & Brugger, P. (2011). When digits help digits: Spatial-numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology*, 2, 1–7.
- Fischer, M. H., Castel, A. D., Dodd, M. D. & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, 6(6), 555–556.
- Fischer, M. H., Mills, R. A. & Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatialnumerical associations. *Brain and Cognition*, 72(3), 333–336.
- Fischer, M. H., Shaki, S. & Cruise, A. (2009). It takes just one word to quash a SNARC. *Experimental Psychology*, 56(5), 361–366.
- Fischer, M. H., Warlop, N., Hill, R. L. & Fias, W. (2004). Oculomotor bias induced by number perception. *Experimental Psychology*, 51(2), 91–97.
- Fueyo, V. & Bushell, D. (1998). Using number line procedures and peer tutoring to improve the mathematics computation of low-performing first graders. *Journal of Applied Behavior Analysis*, 31(3), 417–430.

- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY, US: Springer-Verlag Publishing.
- Fuson, K. C., Richards, J. & Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In C. J. Brainerd (Ed.), *Children's logical and mathematical cognition* (pp. 33–92). Springer New York.
- Galfano, G., Rusconi, E. & Umiltà, C. (2006). Number magnitude orients attention, but not against one's will. *Psychonomic Bulletin & Review*, *13*(5), 869–874.
- Gallistel, C. R. & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1–2), 43–74.
- Gallistel, C. R. & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.
- Galton, F. (1880a). Visualised numerals. *Nature*, *21*, 252–256.
- Galton, F. (1880b). Visualised numerals. *Nature*, *21*, 494–495.
- Galton, F. (1881). Visualised numerals. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *10*, 85–102.
- Gelman, R. (1980). What young children know about numbers. *Educational Psychologist*, *15*(1), 54–68.
- Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In *The psychology of learning and motivation*. (pp. 61–96). San Diego, CA, US: Academic Press.
- Gelman, R. & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA, US: Harvard University Press.
- Gelman, R. & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, *13*(3), 343–359.
- Gelman, R., Meck, E. & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development*, *1*(1), 1–29.
- Gelman, R. & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development*, *46*(1), 167–175.
- Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T. & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, *122*(3), 221–233.
- Gevers, W., Ratinckx, E., De Baene, W. & Fias, W. (2006). Further evidence that the SNARC effect is processed along a dual-route architecture: Evidence from the lateralized readiness potential. *Experimental Psychology*, *53*(1), 58–68.
- Gevers, W., Reynvoet, B. & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–B95.
- Gevers, W., Reynvoet, B. & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *40*(1), 171–172.
- Gevers, W., Verguts, T., Reynvoet, B., Caessens, B. & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology-Human Perception and Performance*, *32*(1), 32–44.
- Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*, *24*(1), 58–87.
- Gielen, I., Brysbaert, M. & Dhondt, A. (1991). The syllable-length effect in number processing is task-dependent. *Perception & Psychophysics*, *50*(5), 449–458.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*,

- 20(1), 1–55.
- Göbel, S. M., Shaki, S. & Fischer, M. H. (2011). The cultural number line: A review of cultural and linguistic influences on the development of number processing. *Journal of Cross-Cultural Psychology*, 42(4), 543–565.
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap Press of Harvard University Press.
- Goldin-Meadow, S., Alibali, M. W. & Breckinridge Church, R. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, 100(2), 279–297.
- Goldin-Meadow, S., Nusbaum, H. C., Garber, P. & Breckinridge Church, R. (1993). Transitions in learning: Evidence for simultaneously activated strategies. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 92–107.
- Goldin-Meadow, S., Wein, D. & Chang, C. (1992). Assessing knowledge through gesture: Using children’s hands to read their minds. *Cognition and Instruction*, 9, 201–219.
- Gordon, D. F. & Desjardins, M. (1995). Evaluation and selection of biases in machine learning. *Machine Learning*, 20(1–2), 5–22.
- Gori, I., Fanello, S. R., Metta, G. & Odone, F. (2012). All gestures you can: A memory game against a humanoid robot. In *Proceedings of the 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*.
- Graham, T. A. (1999). The role of gesture in children’s learning to count. *Journal of Experimental Child Psychology*, 74(4), 333–355.
- Grossberg, S. & Repin, D. V. (2003). A neural model of how the brain represents and compares multi-digit numbers: Spatial and categorical processes. *Neural Networks*, 16(8), 1107–1140.
- Guhe, M., Pease, A., Smail, A., Martinez, M., Schmidt, M., Gust, H., ... Krumnack, U. (2011). A computational account of conceptual blending in basic mathematics. *Cognitive Systems Research*, 12(3-4), 249–265.
- Hafner, V. V. & Kaplan, F. (2005). Learning to interpret pointing gestures: Experiments with four-legged autonomous robots. In S. Wermter, G. Palm & M. Elshaw (Eds.), *Biomimetic neural learning for intelligent robots* (Vol. 3575, pp. 225–234). Berlin Heidelberg: Springer.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hoekstra, J. (1992). Counting with artificial neural networks: An experiment. In I. Aleksander & J. Taylor (Eds.), *Artificial neural networks* (Vol. 2, pp. 1311–1314). Elsevier Science Publishers B.V. (International Conference on Artificial Neural Networks ICANN-92)
- Holmes, K. J. & Lourenco, S. F. (2011). Common spatial organization of number and emotional expression: A mental magnitude line. *Brain and cognition*, 77(2), 315–323.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10), 3088–3092.

- Hostetter, A. B. (2011). When do gestures communicate? a meta-analysis. *Psychological Bulletin*, *137*(2), 297–315.
- Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514.
- Hung, Y.-h., Hung, D. L., Tzeng, O. J. L. & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition*, *106*(3), 1441–1450.
- iCub website*. (2013). Retrieved 3 June 2013, from <http://www.icub.org/>
- Igel, C. & Hüsken, M. (2003). Empirical evaluation of the improved rprop learning algorithms. *Neurocomputing*, *50*(0), 105–123.
- Ishihara, M., Jacquin-Courtois, S., Flory, V., Salemme, R., Imanaka, K. & Rossetti, Y. (2006). Interaction between space and number representations during motor preparation in manual aiming. *Neuropsychologia*, *44*(7), 1009–1016.
- Ito, Y. & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673.
- Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.
- Izard, V. & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. In *Program of the Eighth Annual Conference of the Cognitive Science Society* (pp. 531–546). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kadosh, R. C. & Gertner, L. (2011). Synesthesia: Gluing together time, number, and space. In S. Dehaene & E. M. Brannon (Eds.), *Space, time and number in the brain* (pp. 123–132). San Diego: Academic Press.
- Kaski, S. & Lagus, K. (1996). Comparing self-organizing maps. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen & B. Sendhoff (Eds.), *Artificial neural networks (ICANN 96)* (pp. 809–814). Berlin Heidelberg: Springer.
- Kaufman, E. L., Lord, M. W., Reese, T. W. & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*, 498–525.
- Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B. & Koten, J. W. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *44*(4), 376–385.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (Vol. 25, pp. 207–227). The Hague: Mouton Publishers.
- Kinect for Windows website*. (2013). Retrieved 6 June 2013, from <http://www.microsoft.com/en-us/kinectforwindows/>
- Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549.
- Kiviluoto, K. (1996). Topology preservation in self-organizing maps. In *The 1996 IEEE International Conference on Neural Networks* (Vol. 1, pp. 294–299).
- Kobayashi, T., Hiraki, K. & Hasegawa, T. (2005). Auditory-visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science*, *8*(5), 409–419.

- Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
- Koechlin, E., Dehaene, S. & Mehler, J. (1997). Numerical transformations in five-month-old human infants. *Mathematical Cognition*, 3(2), 89–104.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480.
- Kohonen, T. (2001). *Self-organizing maps* (3rd ed., Vol. 30). Berlin Heidelberg: Springer-Verlag.
- Lakoff, G. & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.
- Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2), 395–438.
- Le Corre, M., Van de Walle, G., Brannon, E. M. & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52(2), 130–169.
- LeCun, Y., Bottou, L., Orr, G. B. & Müller, K.-R. (1998). Efficient backprop. In G. B. Orr & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (Vol. 1524, pp. 9–50). Springer Berlin Heidelberg.
- Leslie, A. M., Gelman, R. & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, 12(6), 213–218.
- Levine, D. S. (2000). *Introduction to neural and cognitive modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Li, X., Gasteiger, J. & Zupan, J. (1993). On the topology distortion in self-organizing feature maps. *Biological Cybernetics*, 70(2), 189–198.
- Lidji, P., Kolinsky, R., Lochy, A. & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1189.
- Link, S. W. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, 34(1), 2–41.
- Lipton, J. S. & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, 14(5), 396–401.
- Lipton, J. S. & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy*, 5(3), 271–290.
- Lungarella, M., Metta, G., Pfeifer, R. & Sandini, G. (2003). Developmental robotics: A survey. *Connection Science*, 15(4), 151–190.
- Ma, Q. & Hirai, Y. (1989). Modeling the acquisition of counting with an associative network. *Biological Cybernetics*, 61(4), 271–278.
- Mandler, G. & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, 111(1), 1–22.
- Mareschal, D. (1998). Developmental cognitive neuroscience and connectionist models of infancy. *Early Development & Parenting*, 7(3), 147–151.
- Marjanović, M., Scassellati, B. & Williamson, M. (1996). Self-taught visually-guided pointing for a humanoid robot. In P. Maes, M. Mataric, J. Meyer, J. Pollack & S. Wilson (Eds.), *From animals to animats 4: Fourth international conference on simulation of adaptive behavior* (pp. 35–44). MIT Press.

- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1–2), 107–157.
- McCloskey, M., Caramazza, A. & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, *4*(2), 171–196.
- McCloskey, M. & Macaruso, P. (1995). Representing and using numerical information. *American Psychologist*, *50*(5), 351–363.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Mechner, F. (1958). Sequential dependencies of the lengths of consecutive response runs. *Journal of the Experimental Analysis of Behavior*, *1*, 229–233.
- Meck, W. H. & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334.
- Metta, G., Fitzpatrick, P. & Natale, L. (2006). YARP: Yet Another Robot Platform. *International Journal of Advanced Robotics Systems*, *3*(1), 43–48.
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., . . . Montesano, L. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, *23*(8–9), 1125–1134.
- Metta, G., Sandini, G., Vernon, D., Natale, L. & Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. In R. Madhavan & E. R. Messina (Eds.), *Proceedings of the 8th Performance Metrics for Intelligent Systems Workshop* (pp. 50–56). ACM.
- Milner, A. D. & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785.
- Mix, K. S., Huttenlocher, J. & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, *128*(2), 278–294.
- Mix, K. S., Levine, S. C. & Huttenlocher, J. (1997). Numerical abstraction in infants: Another look. *Developmental Psychology*, *33*(3), 423–428.
- Moore, D., Benenson, J., Reznick, J. S., Peterson, M. & Kagan, J. (1987). Effect of auditory numerical information on infants' looking behavior: Contradictory evidence. *Developmental Psychology*, *23*(5), 665–670.
- Morse, A. F., de Greeff, J., Belpeame, T. & Cangelosi, A. (2010). Epigenetic robotics architecture (ERA). *Autonomous Mental Development, IEEE Transactions on*, *2*(4), 325–339.
- Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520.
- Moyer, R. S. & Landauer, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*, *1*(3), 167–168.
- Newcombe, N. S. & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: The MIT Press.
- Nieder, A., Freedman, D. J. & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587), 1708–1711.
- Nieder, A. & Miller, E. K. (2003). Coding of cognitive magnitude-compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149–158.
- Nishio, H., Altaf-Ul-Amin, M., Kurokawa, K. & Kanaya, S. (2006). Spherical SOM and Arrangement of Neurons Using Helix on Sphere. *IPSSJ Digital Courier*,

2, 133–137.

- Noël, M., Rousselle, L. & Mussolin, C. (2005). Magnitude representation in children: Its development and dysfunction. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 179–195). New York and Hove: Psychology Press.
- Notebaert, W., Gevers, W., Verguts, T. & Fias, W. (2006). Shared spatial representations for numbers and space: The reversal of the SNARC and the Simon effects. *Journal of Experimental Psychology-Human Perception and Performance*, 32(5), 1197–1206.
- Nuerk, H.-C., Iversen, W. & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 57A(5), 835–863.
- Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, 9(3), 10–20.
- Opfer, J. E. & Furlong, E. E. (2011). How numbers bias preschoolers' spatial search. *Journal of Cross-Cultural Psychology*, 42(4), 682–695.
- Opfer, J. E. & Thompson, C. A. (2006). Even early representations of numerical magnitude are spatially organized: Evidence for a directional magnitude bias in pre-reading preschoolers. In R. Sun & N. Miyaki (Eds.), *Xxviii annual conference of the cognitive science society* (pp. 639–644). Mahwah, NJ: Erlbaum.
- Opfer, J. E., Thompson, C. A. & Furlong, E. E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science*, 13(5), 761–771.
- Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, 91(2), 191–205.
- Pattacini, U., Nori, F., Natale, L., Metta, G. & Sandini, G. (2010). An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1668–1674). Red Hook, NY: IEEE.
- Pepperberg, I. M. (2006). Grey parrot numerical competence: a review. *Animal Cognition*, 9(4), 377–391.
- Perry, M., Breckinridge Church, R. & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, 3(4), 359–400.
- Perry, M., Woolley, J. & Ifcher, J. (1995). Adults' abilities to detect children's readiness to learn. *International Journal of Behavioral Development*, 18(2), 365–381.
- Peterson, S. A. & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science*, 24(1), 93–122.
- Pfeifer, R. & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.
- Piaget, J. (1952). *The child's conception of number*. Oxford, England: W. W. Norton & Co.
- Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, 59(19), 2229–2232.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25.

- Potter, M. C. & Levy, E. I. (1968). Spatial enumeration without counting. *Child Development*, 39(1), 265–272.
- Rajapakse, R. K., Cangelosi, A., Coventry, K. R., Newstead, S. & Bacon, A. (2005a). Connectionist modeling of linguistic quantifiers. In W. Duch, J. Kacprzyk, E. Oja & S. Zadrozny (Eds.), *Artificial neural networks: Formal models and their applications — ICANN 2005* (Vol. 3697, pp. 679–684). Springer Berlin Heidelberg.
- Rajapakse, R. K., Cangelosi, A., Coventry, K. R., Newstead, S. & Bacon, A. (2005b). *Grounding linguistic quantifiers in perception: Experiments on numerosity judgments*. Paper presented at the 2nd Language & Technology Conference, Poznań, Poland.
- Reynvoet, B. & Brysbaert, M. (1999). Single-digit and two-digit arabic numerals address the same semantic number line. *Cognition*, 72(2), 191–201.
- Reynvoet, B. & Brysbaert, M. (2004). Cross-notation number priming investigated at different stimulus onset asynchronies in parity and naming tasks. *Experimental Psychology*, 51(2), 81–90.
- Reynvoet, B., Brysbaert, M. & Fias, W. (2002). Semantic priming in number naming. *The Quarterly Journal of Experimental Psychology Section A*, 55(4), 1127–1139.
- Reynvoet, B., Caessens, B. & Brysbaert, M. (2002). Automatic stimulus-response associations may be semantically mediated. *Psychonomic Bulletin & Review*, 9(1), 107–112.
- Ristic, J., Wright, A. & Kingstone, A. (2006). The number line effect reflects top-down control. *Psychonomic Bulletin & Review*, 13(5), 862–868.
- RobotCub project website*. (2013). Retrieved 3 June 2013, from <http://www.robotcub.org/>
- Rodriguez, P., Wiles, J. & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science*, 11(1), 5–40.
- Roux, F.-E., Boetto, S., Sacko, O., Chollet, F. & Trémoulet, M. (2003). Writing, calculating, and finger recognition in the region of the angular gyrus: a cortical stimulation study of Gerstmann syndrome. *Journal of Neurosurgery*, 99(4), 716–727.
- Ruciński, M., Cangelosi, A. & Belpaeme, T. (2011). An embodied developmental robotic model of interactions between numbers and space. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *33rd annual meeting of the cognitive science society* (pp. 237–242).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 318–362). Cambridge, MA: The MIT Press.
- Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel distributed processing* (Vols. 1–2). Cambridge, MA: The MIT Press.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C. & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, 99(2), 113–129.
- Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N. & Fujimura, K. (2002). The intelligent ASIMO: System overview and integration. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*

- (Vol. 3, pp. 2478–2483). IEEE.
- Salem, M. (2012). *Conceptual motorics — generation and evaluation of communicative robot gesture* (PhD thesis). Bielefeld University.
- Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K. & Joubin, F. (2012). Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, 4(2), 201–217.
- Saxe, G. B. & Kaplan, R. (1981). Gesture in early counting: A developmental analysis. *Perceptual and Motor Skills*, 53(3), 851–854.
- Schaeffel, F. (2007). Processing of information in the human visual system. In A. Hornberg (Ed.), *Handbook of machine vision* (pp. 1–33). Wiley-VCH Verlag GmbH & Co. KGaA.
- Schaeffer, B., Eggleston, V. H. & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, 6(3), 357–379.
- Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., ... Schmidhuber, J. (2010). PyBrain. *Journal of Machine Learning Research*, 11, 743–746.
- Schmitz, A., Maiolino, P., Maggiali, M., Natale, L., Cannata, G. & Metta, G. (2011). Methods and technologies for the implementation of large-scale robot tactile sensors. *Robotics, IEEE Transactions on*, 27(3), 389–400.
- Schwarz, W. & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, 66(4), 651–664.
- Schwarz, W. & Stein, F. (1998). On the temporal dynamics of digit comparison processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1275–1293.
- Sekuler, R. & Mierkiewicz, D. (1977). Children’s judgments of numerical inequality. *Child Development*, 48(2), 630–633.
- Seron, X., Pesenti, M., Noël, M.-P. & Deloche, G. (1992). Images of numbers: or ”when 98 is upper left and 6 sky blue.”. *Cognition*, 44(1–2), 159–196.
- Shaki, S. & Fischer, M. H. (2008). Reading space into numbers — a cross-linguistic comparison of the SNARC effect. *Cognition*, 108(2), 590–599.
- Shaki, S., Fischer, M. H. & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16(2), 328–331.
- Shaki, S. & Gevers, W. (2011). Cultural characteristics dissociate magnitude and ordinal information processing. *Journal of Cross-Cultural Psychology*, 42(4), 639–650.
- Shriki, O., Hansel, D. & Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural computation*, 15(8), 1809–1841.
- Siegler, R. S. & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of Cognitive Skills: The 18th Annual Carnegie Mellon Symposium on Cognition* (pp. 229–294). Hillsdale, NJ: Erlbaum.
- Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A ”non-numerical” account. *Cognitive Development*, 12(3), 349–372.
- Simon, T. J. (1999). The foundations of numerical thinking in a brain without numbers. *Trends in Cognitive Sciences*, 3(10), 363–365.
- Simon, T. J., Hespos, S. J. & Rochat, P. (1995). Do infants understand simple

- arithmetic? A replication of Wynn (1992). *Cognitive Development*, 10(2), 253–269.
- Srinivasan, M. & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition*, 116(2), 217–241.
- Starkey, P. & Cooper, J., Robert G. (1980). Perception of numbers by human infants. *Science*, 210(4473), 1033–1035.
- Starkey, P., Spelke, E. S. & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, 222(4620), 179–181.
- Steels, L. & Brooks, R. A. (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Stoianov, I., Zorzi, M., Becker, S. & Umiltà, C. (2002). Associative arithmetic with boltzmann machines: The role of number representations. In J. R. Dorronsoro (Ed.), *Artificial neural networks – ICANN 2002* (Vol. 2415, pp. 277–283). Springer Berlin Heidelberg.
- Stramandinoli, F., Marocco, D. & Cangelosi, A. (2012). The grounding of higher order concepts in action and language: A cognitive robotics model. *Neural Networks*, 32(0), 165–173.
- Strauss, M. S. & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 52(4), 1146–1152.
- Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L. & Nori, F. (2008). An Open-Source Simulator for Cognitive Robotics Research: The Prototype of the iCub Humanoid Robot Simulator. In R. Madhavan & E. R. Messina (Eds.), *Proceedings of the 8th Performance Metrics for Intelligent Systems Workshop* (pp. 57–61). ACM.
- Tikhanoff, V., Cangelosi, A. & Metta, G. (2011). Integration of speech and action in humanoid robots: iCub simulation experiments. *IEEE Transactions on Autonomous Mental Development*, 3(1), 17–29.
- Trick, L. M. & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently — a limited-capacity preattentive stage in vision. *Psychological Review*, 101(1), 80–102.
- Tversky, B., Kugelmass, S. & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, 23(4), 515–557.
- Uller, C., Carey, S., Huntley-Fenner, G. & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, 14(1), 1–36.
- Ultsch, A. & Siemon, H. P. (1990). Kohonen’s self organizing feature maps for exploratory data analysis. In B. Widrow & B. Angeniol (Eds.), *International neural network conference INNC-90* (pp. 305–308). Dordrecht: Kluwer.
- Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: MIT Press.
- van Dijck, J.-P. & Fias, W. (2011). A working memory account for spatialnumerical associations. *Cognition*, 119(1), 114–119.
- van Dijck, J.-P., Gevers, W., Lafosse, C., Doricchi, F. & Fias, W. (2011). Non-spatial neglect for the mental number line. *Neuropsychologia*, 49(9), 2570–2583.
- van Galen, M. S. & Reitsma, P. (2008). Developing access to number magnitude: A

- study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology*, *101*(2), 99–113.
- Van Loosbroek, E. & Smitsman, A. W. (1990). Visual perception of numerosity in infancy. *Developmental Psychology*, *26*(6), 916–922.
- van Opstal, F., Fias, W., Peigneux, P. & Verguts, T. (2009). The neural representation of extensively trained ordered sequences. *Neuroimage*, *47*(1), 367–375.
- Verguts, T. & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504.
- Verguts, T. & Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Philosophical Psychology*, *21*(4), 539–554.
- Verguts, T., Fias, W. & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66.
- Verguts, T. & van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review*, *12*(5), 925–930.
- Vernon, D., von Hofsten, C. & Fadiga, L. (2010). *A roadmap for cognitive development in humanoid robots*. Berlin: Springer-Verlag.
- Von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology*, *18*(3), 450–461.
- Vuilleumier, P., Ortigue, S. & Brugger, P. (2004). The number space and neglect. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *40*(2), 399–410.
- Wang, H., Johnson, T. R., Sun, Y. & Zhang, J. (2005). Object location memory: The interplay of multiple representations. *Memory & Cognition*, *33*(7), 1147–1159.
- Washburn, D. A. & Rumbaugh, D. M. (1991). Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science*, *2*(3), 190–193.
- Whalen, J., Gallistel, C. R. & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137.
- Widrow, B. & Lehr, M. A. (1990). 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, *78*(9), 1415–1442.
- Wilson, H. R. & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*(1), 1–24.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.
- Wood, G., Willmes, K., Nuerk, H.-C. & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science*, *50*(4), 489–525.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193.
- Wynn, K. (1992a). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750.
- Wynn, K. (1992b). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251.
- Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, *7*(3), 164–169.
- Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, *2*(8), 296–303.

- Xu, F. & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, 30(2), 111–153.
- Xu, F. & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1-B11.
- Xu, F., Spelke, E. S. & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8(1), 88–101.
- Yamashita, Y. & Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment. *Plos Computational Biology*, 4(11).
- Zbrodoff, N. J. & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331–345). New York and Hove: Psychology Press.
- Zebian, S. (2005). Linkages between Number Concepts, Spatial Thinking, and Directionality of Writing: The SNARC Effect and the REVERSE SNARC Effect in English and Arabic Monoliterates, Bilinguals, and Illiterate Arabic Speakers. *Journal of Cognition and Culture*, 5(1–2), 165–190.
- Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature*, 284(5755), 412–418.
- Zorzi, M. & Butterworth, B. (1997). On the representation of number concepts. In M. Shafto & P. Langley (Eds.), *19th Annual Conference of the Cognitive Science Society* (p. 1098). Mahwah, NJ: Cognitive Science Society.
- Zorzi, M. & Butterworth, B. (1999). A computational model of number comparison. In M. Hahn & S. Stoness (Eds.), *21st Annual Meeting of the Cognitive Science Society* (pp. 778–783). Mahwah, NJ: Cognitive Science Society.
- Zorzi, M., Priftis, K. & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature*, 417(6885), 138–139.
- Zorzi, M., Stoianov, I. & Umiltà, C. (2005). Computational modeling of numerical cognition. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 67–83). New York and Hove: Psychology Press.

Publications

On the remaining pages the following publications are reproduced:

Ruciński, M., Cangelosi, A. & Belpaeme, T. (2011). An embodied developmental robotic model of interactions between numbers and space. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *33rd annual meeting of the cognitive science society* (pp. 237–242).

Stramandinoli, F., Ruciński, M., Znajdek, J., Rohlfing, K. J. & Cangelosi, A. (2011). From sensorimotor knowledge to abstract symbolic representations. *Procedia Computer Science*, 7, 269–271.

Ruciński, M., Cangelosi, A. & Belpaeme, T. (2012). Robotic model of the contribution of gesture to learning to count. In *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob 2012)* (pp. 1–6). © 2012 IEEE. Reprinted with permission.

An Embodied Developmental Robotic Model of Interactions between Numbers and Space

Marek Ruciński (marek.rucinski@plymouth.ac.uk)
Angelo Cangelosi (angelo.cangelosi@plymouth.ac.uk)
Tony Belpaeme (tony.belpaeme@plymouth.ac.uk)

Centre for Robotics and Neural Systems, University of Plymouth
Plymouth, Devon, PL48AA, UK

Abstract

In this paper we describe an embodied developmental model of the interactions between the neural representations of numbers and space in the humanoid robot iCub. We show how a simple developmental process that mimics real-world cultural biases leads to the emergence of certain properties of the number and space representation system that enable the robot to reproduce well-known experimental phenomena. We demonstrate the validity of the proposed approach by showing that it leads to the reproduction of three psychological phenomena connected with number processing, namely size and distance effects, the SNARC effect and the Posner-SNARC effect. **Keywords:** mathematical cognition; developmental cognitive robotics; computational modeling; size effect; distance effect; SNARC effect; Posner-SNARC effect;

Introduction

Perceiving numbers and quantities is one of the most basic perceptual skills of humans and animals (Dehaene, 1997). Given the pure and abstract character of the number concept as perceived by humans, it is no surprise that many in cognitive science pursue a better understanding of how such a peculiar concept could have emerged, how it is represented and processed, and how it relates to other processes that take place in the brain. These efforts, which can be put together under the common label of mathematical cognition, constitute a branch of science that has been gaining more and more momentum during the past few decades (Dehaene & Brannon, 2010).

Computational modeling is an important tool used in the study of mathematical cognition to understand the principles of number processing in the brain. Based on observations from experimental psychological studies as well as hints obtained through various brain imaging techniques, computer models of number representation and processing are constructed and evaluated on the basis of how well their properties match those of the biological cognitive systems. Analysis of the computer models helps us to understand how biological systems work at the algorithmic level, which in turn is necessary to understand their neural implementations.

In this paper we present an embodied developmental cognitive robotic model of interactions between number and space. In the following paragraphs we provide a short review of previous computational models of numerosity representation and processing, focusing on those most relevant to the work presented herein.

An influential connectionist model of number representation and processing has been described by Dehaene and

Changeux (1993). The system consisted of a 1-dimensional visual retina, a location and normalization cluster, a summation coding layer and a place coding layer. The output from the place coding layer was used in a “same-different” comparison and “larger-smaller” comparison tasks. The system was designed to model perception and processing of non-symbolic stimuli (e.g. a cardinality of a set of items perceived visually). One of the most interesting findings was the demonstration of how the described system can autonomously “discover” the larger-smaller relation based solely on unsupervised experimentation with addition and subtraction.

One of the first models of number representation based on recurrent artificial neural networks was proposed by Rodriguez, Wiles, and Elman (1999). Here, supervised learning techniques were used to teach a simple recurrent neural network to perform a task, in which counting was required in order to succeed. In successful networks, neurons formed a special case of a discrete-time dynamical system in which numerosity was coded in the dynamical properties of trajectories realized by hidden layer units. This complex solution, radically different from traditionally used coding methods, has been obtained despite a small amount of inductive bias in the training process.

Ahmad, Casey, and Bale (2002) presented a rather complex system aimed at modeling two manifestations of numerical abilities: subitizing and counting. Their system consisted of two networks, each specifically designed to perform one of these tasks, composed of several modules playing different roles and trained separately using various machine learning techniques. Implementation of the model delivered interesting results especially in the domain of counting (which, being a more complex task with a temporal structure, has been more rarely tackled in the literature than instant comprehension of numerosity), where counting error patterns similar to those observed in children were obtained.

A relatively consistent path of increasingly complex modeling of different aspects of human mathematical cognition can be found in a series of papers by Verguts and collaborators (Verguts & Fias, 2004; Verguts, Fias, & Stevens, 2005; Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006; Chen & Verguts, 2010). The first model focused on how simple number coding methods believed to be employed in the brain (summation and place coding) can emerge as the result of an unsupervised learning process, thus showing that

such systems do not have to be innate as suggested in earlier research. Building on a place-coding system with linear scaling and constant variability as the core representation of numerosity, Verguts et al. (2005) shifted the responsibility for size and distance effects from number representation to later processing stages. It was demonstrated that this leads to results consistent with experimental data. These are characterized by symmetrical priming patterns and no size effect in naming and parity tasks, combined with the presence of both size and distance effects in the comparison task. This is allegedly not possible to obtain using numerosity representations with compressed scaling and/or increasing variability, used in earlier models. An important step has been achieved by Gevers et al. (2006), where experimental phenomenon more complex than size and distance effects, namely the SNARC effect (Spatial-Numerical Association of Response Codes, (Dehaene, Bossini, & Giraux, 1993)) were modeled. The model used a dual-route architecture to explain the phenomenon, combining findings from previous computational models and other studies aiming at explaining spatial congruency effects. The simulations were compared to experimental data, predictions were made about the shape of the SNARC effect in a certain category of tasks, and these were confirmed experimentally.

Finally, the model of Gevers et al. (2006) was further extended in a recent paper by Chen and Verguts (2010), in which a representation of space was introduced instead of an “automatic pathway” present in the previous model. Chen and Verguts (2010) added a module corresponding to a “human homologue of lateral intra-parietal area in macaque monkeys”, a saliency map related to the visual field, consisting of two parts characterized by contra-lateral spatial neuronal gradients. These gradients were identified as the crucial property of the model which allowed for reproduction of a number of psychological experiments, including those involving patients suffering from certain lesions.

The model we present in this paper extends the work of Chen and Verguts (2010) by addressing two drawbacks with their model. First, as it is the case with all mathematical cognition models published to-date, the system does not take directly into account any aspects of embodiment. According to current trends in cognitive science it is not possible to understand the brain in separation from the body in which it is embedded and from the environment in which it develops. In line with this, when formulating our model we considered any relevant constraints imposed by the target body (that of a humanoid robot), and designed the developmental process accordingly. Secondly, the most important phenomenon investigated by Chen and Verguts (2010), that is associations between numbers and space, have been modeled in their paper as hand-wired connections, despite extensive evidence cited by themselves that most probably it is the “environmental correlation between symbolic numbers and physical space” that creates this association in the brain. In this paper we show how necessary patterns of connections can indeed

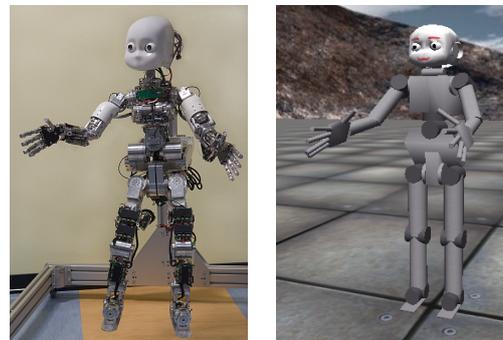


Figure 1: iCub, the humanoid robot used in modeling.

emerge from a simple developmental process.

The following sections of the paper are organized as follows. First we introduce the robotic platform that has been employed in this modeling study. Then we present the architecture of our model and the process of its development. Next we demonstrate the validity of our model by showing that it is able to reproduce three phenomena in which interactions between numbers and space manifest themselves. We finish the article by drawing conclusions from the experiments and emphasizing the capability of the embodied robotic approach to be used in the modeling of mathematical cognition.

Model Description

iCub, the Humanoid Robot Platform

The model described in this paper has been designed to operate in a simulated model of the humanoid robot iCub (Metta et al., 2010). The robot itself (figure 1), is an open-source design developed recently as a benchmark platform for cognitive robotics experiments. The anatomy of the robot is intended to resemble that of a 3.5 years old human child and has a total of 53 degrees of freedom, 20 of which were used in the experiments described here (6 for head and eyes, and 7 for each of the two arms). iCub is equipped with devices which allow it for visual, auditory, tactile and proprioceptive perception. Robot software includes the iCub simulator (Tikhonoff, Cangelosi, & Metta, in press), a tool for robotic simulation experiments without the use of the physical robot. In research described in this paper only the simulated robot has been used.

Model Architecture

The architecture of the model (see figure 2) builds on results of the modeling experiments described above, as well as those of Caligiore, Borghi, Parisi, and Baldassarre (2010), where the authors formulated a general embodied model of compatibility effects focusing on motor affordances and goals. The processing of information in our model is split into two neural pathways: “ventral”, responsible for processing the identity of objects as well as task-dependent decision making

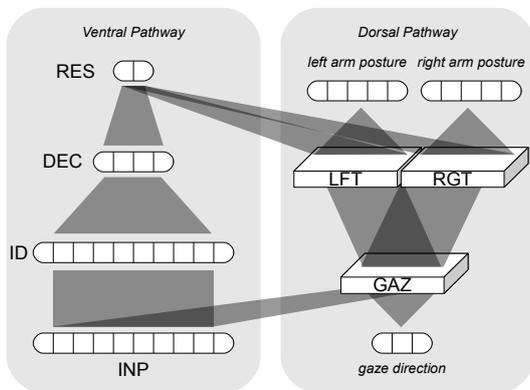


Figure 2: Architecture of the model.

and language processing, and “dorsal”, involved in processing of spatial information about the locations and shapes of objects and sensorimotor transformations which provide on-line support for visually guided motor actions (please refer to (Caligiore et al., 2010) for an extensive discussion of motivations for such a division).

“Ventral” Pathway: Decision Making and Language Processing The “ventral” pathway is modeled in a very similar way to components of the (Chen & Verguts, 2010) model. It consists of: 1) a *symbolic input* INP which codes for the number, using place coding (same remarks about the irrelevance of the spatial arrangement of neurons as those raised in the original paper apply here); 2) a “mental number line” ID which codes for number *identity* (the meaning of the symbol) with linear scaling and constant variability; 3) a *decision layer* DEC executing each of the considered tasks, that is number comparison and parity judgment; and 4) a *response layer* RES, integrating information from both pathways and responsible for the final selection of the motor response. Similar to the practice used in (Chen & Verguts, 2010), for simplicity of implementation the actual structure of the “ventral” pathway, especially its decision layer, was adapted depending on the task to be performed, by removing components irrelevant to the task at hand. Likewise, for the number comparison task which requires more than one number to be processed at the same time, short-term memory was implemented by duplicating necessary layers (namely INP and ID). The layers were composed of the following numbers of neurons: INP and ID: 15, DEC: 4 (2 for each task), RES: 2.

“Dorsal” Pathway: Spatial Coding and Transformations The “dorsal” pathway is composed of a number of neuronal maps which code for the spatial locations of objects in the robot peripersonal working space using different frames of reference (Wang, Johnson, & Zhang, 2001): one associated with the gaze direction (GAZ), and two for each of the robot’s arms: left (LFT) and right (RGT). These maps are

implemented as 49-cell (7 by 7) 2-dimensional Kohonen Self-Organizing Maps (SOMs) with cells arranged in a hexagonal pattern. Input to the GAZ map comes from the 3-dimensional proprioceptive vector representing the robot gaze direction (azimuth, elevation and vergence) and input to each arm position map consists of a 7-dimensional proprioceptive vector representing the position of the relevant arm joints: shoulder pitch, roll and yaw, elbow angle and wrist pronosupination, pitch and yaw. The GAZ map is linked to both arm maps: this implements the transformation of spatial coordinates between frames of reference corresponding to these body parts (so that a position in the visual field can be translated into an arm posture corresponding to reaching to this position and vice-versa). It is important to note that this is the part of the model where the embodied approach to modeling is implemented, and where the crucial difference between our and all previous quoted models lies. This point is elaborated in the Discussion section.

Developmental Learning of the Robot

The modeling of the developmental learning process is organized around a number of sequential phases corresponding to different stages of development of a human child. First, spatial representations for sight and motor affordances have to be built and correspondences between them established. Later, the child can learn number words and their meaning. Usually in late preschool years, children learn to count. More or less at the same stage the child may be taught to perform simple numerical tasks such as number magnitude comparison or parity judgment. All these stages are reflected in our model.

Building Spatial Representations and Transformations

In order to build the gaze and arm space maps, the robot performs a process equivalent to *motor babbling* (Von Hofsten, 1982), in which a child refines its internal visual and motor space representations by performing random movements with arms while observing its hands, reaching for toys in its visual field, etc. This enables the child to perform tasks such as visually guided reaching later in life. This stage of development was implemented in the robot by selecting 90 points uniformly distributed on what has been assumed to be the robot’s operational space (a part of a sphere in front of the robot with 0.65m radius, centered between robot’s shoulder joints, spanning $\pm 30^\circ$ of elevation and $\pm 45^\circ$ in azimuth). These points served as target locations for directing gaze and moving both arms of the robot using inverse kinematic modules. After a trial in which the robot reached a random position, the resulting gaze and arm postures were read from proprioceptive inputs and stored. Between each trial, the head and arms of the robot were moved to the rest position in order to eliminate any influence of the sequence in which the points have been presented on the head and arm posture at the end of the motion. These data were used to train the three SOMs using the traditional unsupervised learning algorithm. In order to reflect the asymmetry between reachable space for the left and right arm (some areas reachable by the right arm cannot be reached by

the left arm and vice versa), only 2/3 of the extreme points corresponding to an arm were used when building a spatial map for this arm (e.g. leftmost 2/3 of all points for the left arm). Learning parameters were adjusted manually based on the observation of the learning process and analysis of how well resulting networks span target spaces.

Transformations between the visual spatial map GAZ and the maps of reachable space LFT and RGT, implemented as connections between the maps, were trained using the classical Hebbian learning rule. In a process similar to motor babbling, gaze and the appropriate arm were directed toward the same point and resulting co-activations in already developed spatial maps were used to establish links between them.

Learning Number Words and Their Meaning This stage of learning corresponds to establishing links between number words, modeled as activations in the INP layer, and number meaning, being activations in the ID layer. In the model described here links between INP and ID layers were preset manually implementing place coding with linear scaling and constant variability (as in (Chen & Verguts, 2010) and previous models). However, Verguts and Fias (2004) showed that such pattern of connections can arise from a simple supervised learning process.

Learning to Count The goal of this stage is to model the cultural biases that result in an internal association of “small” numbers with the left side of space and “large” numbers with the right side, since this is believed to be the cause of SNARC and similar effects. As an example of these biases we considered a tendency of children to count objects from left to right, which may be associated with the fact that European culture is characterized by left-to-right reading direction (Dehaene, 1997). In order to model the process of learning to count, the robot was exposed to an appropriate sequence of number words (fed to the INP layer of the model network), while at the same time robot’s gaze was directed toward a specific location in space (via the input to the GAZ spatial map). These spatial locations were generated in such a way that their horizontal coordinates correlated with number magnitude (low number presented on the left, large numbers on the right) with a certain amount of Gaussian noise. Vertical coordinates were chosen to uniformly span the represented space. While the robot is exposed to this process, Hebbian learning establishes links between number word and stimuli location in the visual field.

Learning Comparison and Parity Tasks Finally, the model is trained to perform target tasks, that is number comparison and parity judgment, which corresponds to establishing appropriate links between the ID layer and neurons in the DEC layer. This process, extensively described in (Verguts et al., 2005), involves supervised learning using the Widrow-Hoff Delta learning rule after all activations in the network reach stable states. In our model we used weight values from our own reproduction of the experiments described in

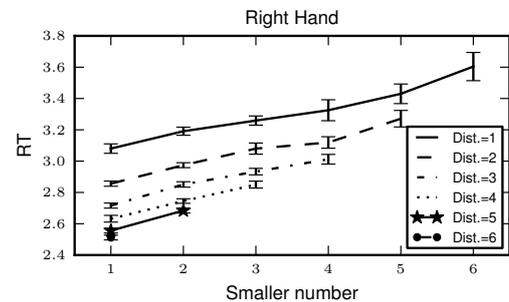


Figure 3: Simulation of the size and distance effects in the number comparison task. On all RT charts error bars show $\pm 2SEM$.

the cited paper.

Simulation Results

In order to demonstrate the validity of our model we tested it by simulating three selected tasks which have been used previously to evaluate models by other authors (Chen & Verguts, 2010). In this section we present a brief summary of the results. All of the tasks involved measuring response times (RT) of the model. These were obtained by assuming that a response is given when activity in one of the two response nodes exceeded an assumed response threshold (0.5 for experiments 1 and 2 and 0.8 for experiment 3). We report RTs aggregated over 10 independent instantiations of the model¹.

Experiment 1: Size and Distance Effects

Size and distance effects are two of the most common findings from experimental mathematical cognition studies. They are present in many tasks, but in the context of number comparison they mean that it is more difficult to compare larger numbers (size effect) and numbers which are closer to each other (distance effect). This should be evident in RTs growing with number magnitude and with decreased distance between numbers being compared. RTs obtained from simulating the experiment in our model are reported in figure 3. Response times were measured for all pairs of numbers from 1 to 7. We report results for the right hand response only (results for the left hand were similar). Clearly both size and distance effects are present in the model. Sources of the size and distance effects in our model are the same as in the model by Chen and Verguts (2010), namely monotonic and compressive patterns of weights between ID and DEC layers.

¹In contrast to cited authors we had no access to numerical data from relevant psychological experiments, thus we were unable to perform linear regression over these data. This does not invalidate any of our results (only additional linear scaling of RTs is performed), but must be kept in mind when comparing charts from the respective papers.

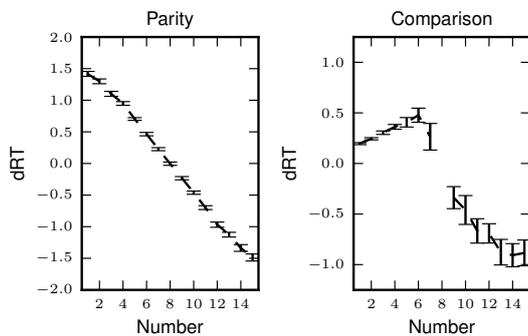


Figure 4: Simulation of the SNARC effect in parity judgment and magnitude comparison tasks.

Experiment 2: SNARC effect

SNARC effect is more directly related to interactions between number and space than size and distance effects. Using a similar procedure as in (Chen & Verguts, 2010), we report RTs obtained by our model in parity judgment and number comparison tasks. Here, the difference between right hand and left hand RTs for the same number in both congruent and incongruent condition is reported. The SNARC effect should manifest itself in a negative slope on such a chart. Results of our simulations are presented in figure 4. Presence of the SNARC effect is evident in both tasks. The source of this effect in our model requires further explanation.

Quoting relevant neuroscientific research, Chen and Verguts (2010) explain sources of the SNARC effect as the result of "an initial dip toward the wrong response hand in SNARC-incongruent conditions evident in recordings of the lateralized readiness potentials in the motor cortex". Accordingly, in our model the presentation of a number word leads to an automatic activation of the relevant parts of the visual space representation, due to links established during model development (more precisely, during learning to count) – left part for small numbers, and right part for large ones. Visual space representations in turn are linked to both motor maps, although not symmetrically. As outlined above in the description of the model development, some parts of the visual space that can be reached by the right arm cannot be reached by the left arm, and vice versa. As a consequence, when transformation from the visual space map to arm maps occurs, both arm-related representations will be activated to a similar degree only for the areas in the center of the visual map. For the areas placed to the sides of the visual space, the map associated with one arm will be activated more strongly than the other, as it over-represents that side of space (this is a natural consequence of the robot morphology). Because there is a significant overlap between represented areas, the effect is not sudden, but connections between visual and motor maps form a gradient from left to right – links to the left arm map become weaker, while those to the right become stronger.

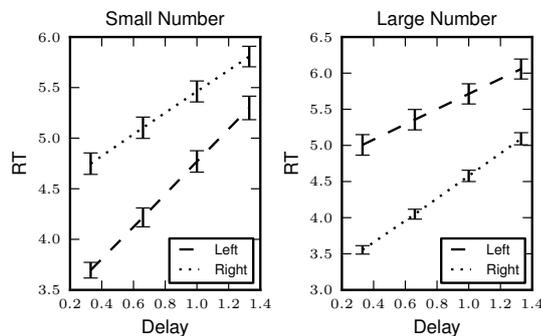


Figure 5: Simulation of the Posner-SNARC effect.

Thus, for instance when a small number is presented, internal connections lead to stronger automatic activation of the representations linked with the left arm than those of the right arm, which causes the SNARC effect. In contrast to (Chen & Verguts, 2010), in our model this particular pattern of connections is not hand-wired, but emerges as a consequence of the robot morphology during the development process. We hypothesize that the presence of such neuronal gradients in the human brain referred to by (Chen & Verguts, 2010) may be ascribed to similar factors.

Experiment 3: Posner-SNARC effect

The Posner-SNARC effect is another manifestation of the connection between numbers and space, placed within the attention cuing paradigm (Fischer, Castel, Dodd, & Pratt, 2003). A small or large number presented at the fixation point acts as a cue and directs attention of the participant toward the left or right side of space, affecting the time needed to detect an object appearing in the visual field after a certain delay. The effect results in faster detection of the target on the left when a small number is presented as a cue, and on the right for large numbers, even though throughout the experiment numbers are not predictive of target locations. Simulated response times obtained from model are shown in figure 5. The effect is visible on the charts in shorter RTs for the target presented on the left for a small number as a cue, and on the right for a large number as a cue.

Discussion

In the paper we have presented an embodied developmental robotic model of interactions between numbers and space. We have described the model architecture as well as the associated developmental process. By simulating three well-known experiments we have demonstrated the validity of our approach, showing that after development is completed, our model exhibits the most important properties of the human mathematical cognition system. In this final section of the article we discuss the differences between our approach and those of authors of earlier works, thus highlighting the benefits which embodied robot simulations bring to cognitive

modeling in general.

As described above, the crucial difference between our modeling approach and previous literature models is the aspect of embodiment. The robot we use in our experiments has one head and two arms, thus three separate spatial maps for each of these body parts are developed in our cognitive model. The robot proprioceptively perceives its gaze direction and arm positions using specific degrees of freedom, and as a consequence the maps in our model have to be implemented to span this specific number of dimensions. Finally, these maps are *real* spatial maps, in which activations correspond to specific positions of a material limb and vice versa. Thus such an embodied approach may greatly help to reduce arbitrariness of the model. Taking as an example the system described by Chen and Verguts (2010), “space representation” has been implemented there as an arbitrary network of connections, hand-wired in such a way so that it exhibits properties suggested by neuroscientific data. Although this allowed for a successful reproduction of a good number of experiments, the traditional connectionist approach remained inconclusive regarding how to answer the questions of *why* such a pattern of connections is present and *how* it comes into being. Supplementing the previous modeling achievements with the embodied approach and replacing previously arbitrary parts of the model with elements which have direct material interpretation allowed us to formulate hypotheses to answer these questions.

The importance of the embodied approach to cognitive modeling increases together with the level of complexity of the processes being modeled, and with the degree to which motor representations and actions are involved. In the context of mathematical cognition one may recall experiments such as physical line bisection (where the participant is asked to point at the middle of a line presented on board in front of him) or investigation of the role of finger counting habits (Fischer, 2008) or that of gesture in learning to count (Andres, Seron, & Olivier, 2007) to name just a few. While some researchers already attempted to model the former task with a purely connectionist model (Chen & Verguts, 2010), the embodied robotic approach is more suited to tackle such problems from the developmental perspective.

Results presented in this paper are part of work in progress. After connecting the model with the real iCub robot instead of its virtual equivalent, we plan to employ it to tackle issues in mathematical cognition directly involving motor representations and actions, with a special focus on the relations between gesturing and learning to count.

Acknowledgments

This research has been supported by the EU project RobotDoC (235065) from the FP7 Marie Curie Actions ITN.

References

Ahmad, K., Casey, M., & Bale, T. (2002). Connectionist simulation of quantification skills. *Connection Science*, *14*(3), 165-201.

- Andres, M., Seron, X., & Olivier, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience*, *19*(4), 563-576.
- Caligiore, D., Borghi, A. M., Parisi, D., & Baldassarre, G. (2010). TRoPICALS: A computational embodied neuroscience model of compatibility effects. *Psychological Review*, *117*(4), 1188 - 1228.
- Chen, Q., & Verguts, T. (2010). Beyond the mental number line: A neural network model of number-space interactions. *Cognitive Psychology*, *60*(3), 218-240.
- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371 - 396.
- Dehaene, S., & Brannon, E. M. (2010). Space, time, and number: a Kantian research program. *Trends in Cognitive Sciences*, *14*(12), 517-519.
- Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *J. Cognitive Neuroscience*, *5*(4), 390-407.
- Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex*, *44*(4), 386-392.
- Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, *6*(6), 555-556.
- Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology-Human Perception and Performance*, *32*(1), 32-44.
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, *23*(8-9), 1125 - 1134.
- Rodriguez, P., Wiles, J., & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science*, *11*(1), 5 - 40.
- Tikhonoff, V., Cangelosi, A., & Metta, G. (in press). Language understanding in humanoid robots: iCub simulation experiments. *IEEE Transactions on Autonomous Mental Development*.
- Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493-1504.
- Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66.
- Von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology*, *18*(3), 450-461.
- Wang, H., Johnson, T. R., & Zhang, J. (2001). The mind's views of space. In *Proceedings of the Third International Conference on Cognitive Science* (pp. 191-198).



The European Future Technologies Conference and Exhibition 2011
From Sensorimotor Knowledge to Abstract Symbolic
Representations[☆]

Francesca Stramadinoli^a, Marek Ruciński^{a,*},
Joanna Znajdek^b, Katharina J. Rohlfing^b, Angelo Cangelosi^a

^a Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, PL48AA Plymouth, United Kingdom

^b Cognitive Interaction Technology Center of Excellence, Bielefeld University, Universitätsstraße 2123, 33615 Bielefeld, Germany

Abstract

We present two cognitive robotic experiments looking at different aspects of relations between symbolic representations and sensorimotor knowledge.

© Selection and peer-review under responsibility of FET11 conference organizers and published by Elsevier B.V.

Keywords: Developmental cognitive robotics; Mathematical cognition; Symbol grounding

Developmental cognitive robotics permits the modeling of different brain and behavioral processes that take place during child development. In contrast to purely computational modeling methods, the principal advantage of the robotic approach is that it enables inherent inclusion of different aspects of sensorimotor control and representation in the model, consistent with the embodied view of cognition. Traditional cognitive robotic modeling research puts however a lot of emphasis on the processes connected with motor behavior itself. We would like to extend this view by looking at the relations between motor actions and abstract symbol manipulation capabilities.

The development of symbol manipulation capabilities in children such as productive language use is preceded by the establishment of a variety of both verbal and non-verbal communication routines with their caregivers. Such routines are grounded in multi-modal interaction practices that are temporally coordinated and contingent with the interlocutor's feedback. E.g. Nomikou and Rohlfing [1] found that when speaking with their 3 month old infants, mothers vocalize in a tight temporal relationship with action over a considerable part of the overall interaction time, thereby making the vocal signal both perceivable and tangible to the infants. In later practices, adults use combination of pointing, showing and words to describe an action or an object and highlight its specific features. The child acquires the symbolic meaning of these words and actions by a frequent observation of the parents and reception of their feedback in response to their own actions. All these observations suggest the existence of a strong link between the sensorimotor knowledge and the abstract symbol manipulation abilities which has been the topic of two cognitive robotic experiments described below.

The first issue addressed in our experiments is the development of linguistic skills. Language capabilities are one of the most powerful tools of an agent for understanding situations and interacting with other agents in the environment. In the framework of cognitive science, psychological experiments that focus on the relationship between language

[☆] This research has been supported by the EU project RobotDoC (235065) from the FP7 Marie Curie Actions ITN.

* Corresponding author.

E-mail addresses: francesca.stramandinoli@plymouth.ac.uk (F. Stramadinoli), marek.rucinski@plymouth.ac.uk (M. Ruciński), jznajdek@techfak.uni-bielefeld.de (J. Znajdek), rohlfing@techfak.uni-bielefeld.de (K.J. Rohlfing), angelo.cangelosi@plymouth.ac.uk (A. Cangelosi).

development and other cognitive capabilities (e.g. perception, action) have been presented. According to the results of these experiments, the development of linguistic skills requires different cognitive processes working together; nevertheless, other models proposed in the field of language learning systems mainly focus on the idea that language is an independent and autonomous capability of agents. In our experiments, we propose a model based on Artificial Neural Networks (ANNs) for symbols manipulation that provides a useful tool for investigating and testing embodied theories of language learning. Experiments take inspiration from the model proposed by Cangelosi and Riga [2], in which a simulated robot was trained first by using the mechanism of the *direct grounding* for learning a set of action primitives and their corresponding name and then by the mechanism of the *grounding transfer* by which the grounding of basic words is transferred to higher-order words via linguistic description.

Simulation experiments have been developed on a software environment for the iCub robot. A set of words, that express general actions with a sensorimotor component, were first taught to the simulated robot through direct grounding mechanism; subsequently, by combining words grounded in sensorimotor experience, the simulated robot acquired more abstract concepts. In particular, the training of the robot consisted of three incremental stages: (i) Basic Grounding (BG) phase for learning to perform a set of basic action primitives and their corresponding names (e.g. GRASP, STOP, SMILE), (ii) Higher-order Grounding 1 (HG1) stage when the robot, via linguistic description, acquires higher-order words by combining basic *action primitives* (e.g. KEEP is GRASP and STOP) and (iii) Higherorder Grounding 2 (HG2) stage during which the robot learns *high-level behaviors* through the combination of action primitives and *higher-order words* (e.g. ACCEPT is KEEP and SMILE and STOP). Simulations results demonstrate that higher-order symbolic representations and behaviors can be indirectly-grounded in basic action primitives directly-grounded in sensorimotor experience. This model is being extended to test other embodied cognition theories of language learning such as the Action-sentence Compatibility Effect.

The second experiment focuses on one of the most established psychological phenomena that suggests the existence of a link between symbolic numbers and motor space representations in the brain, that is so-called SNARC effect (Spatial-Numerical Association of Response Codes). The effect means that when responding to a small number, the reaction time for a left hand response is faster than for the right hand, and conversely for a large number it takes longer to respond with a left hand than with a right hand. The existence of number-space associations is also supported by data from other disciplines like neuroscience, studies of patients with lesions and computational modeling [3].

Building on the results from previous modeling experiments, we formulated an embodied developmental robotic model of interactions between numbers and space. It is composed of a *ventral* pathway, responsible for symbolic tasks, i.e. language processing, coding of the identity of objects as well as task-dependent decision making, and a *dorsal* pathway, involved in processing spatial data about locations of objects in different frames of reference and allowing for appropriate transformations. The latter pathway implements the aspect of embodiment, as its elements are designed to map various parts of the iCub robot, namely positions of its two arms and the gaze direction. We also proposed a developmental process for the model that resembles that of a human child. First, in a process alike to *motor babbling*, space representations in the dorsal pathway and links between them are constructed. Next, the model is taught number words and their meaning. Then, the robot is taught to count, in a way resembling the real-life cultural biases: even though the objects being counted are placed randomly in the visual field of the robot, the process always proceeds from left to right. Finally, the robot is taught to perform tasks like parity judgment and magnitude comparison that enables assessment of various effects of embodiment. The model developed using the described process successfully reproduces major effects known from psychological studies, such as the already mentioned SNARC effect, as well as the Posner-SNARC effect, and size and distance effects in number comparison. This suggests that numbers-space associations may be the result of cultural biases present in the environment during the course of child development. In future we plan to extend the model to allow for investigation of the importance of gesture in learning to count.

In our experiments we looked at two aspects of relations between sensorimotor knowledge and symbol manipulation capabilities. Apart from delivering new results in respective areas of cognitive science, both studies demonstrate the potential of cognitive robotic models for the investigation of the human cognitive development. Embodied character of such modeling approach allows for a more accurate analysis of the corresponding biological processes.

References

- [1] I. Nomikou, K. Rohlfing, Language does something: Body action and language in maternal input to 3-month-olds, IEEE Transactions on Autonomous Mental Development.

- [2] A. Cangelosi, T. Riga, An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots, *Cognitive Science* 30 (4) (2006) 673–689.
- [3] Q. Chen, T. Verguts, Beyond the mental number line: A neural network model of number-space interactions, *Cognitive Psychology* 60 (3) (2010) 218–240.

Robotic Model of the Contribution of Gesture to Learning to Count

Marek Ruciński*, Angelo Cangelosi, Tony Belpaeme
Centre for Robotics and Neural Systems, Plymouth University
Drake Circus, PL4 8AA, Plymouth, United Kingdom
Email: marek.rucinski,angelo.cangelosi,tony.belpaeme@plymouth.ac.uk
*Telephone: +44 (0) 17525 84908

Abstract—In this paper a robotic connectionist model of the contribution of gesture to learning to count is presented. By formulating a recurrent artificial neural network model of the phenomenon and assessing its performance without and with gesture it is demonstrated that the proprioceptive signal connected with gesture carries information which may be exploited when learning to count. The behaviour of the model is similar to that of human children in terms of the effect of gesture and the size of the counted set, although the detailed patterns of errors made by the model and human children are different.

I. INTRODUCTION

It is widely accepted that gestures like pointing, touching or moving items when counting are an integral part of the development of children’s number knowledge [1]. Children use such gestures spontaneously and many studies have confirmed that it facilitates counting accuracy [1]–[8]. As number knowledge can be regarded as a major example of abstract thought, it is not surprising that counting skill has been drawing the attention of psychologists for a long time [9]. As a result, a lot of valuable experimental data on the contribution of gesture to learning to count have been gathered. For instance it is known that prevention of pointing disrupts the counting procedure; in such situations a child usually emits an indefinite stream of number words or does not count at all [2]. In addition, evidence has been provided for the importance of the actual physical contact with counted objects; counting items behind a transparent cover proves to be more difficult for children than when they are allowed to touch the objects being counted [3], [5]. It is also known that gesture plays a developmental role: it is particularly helpful for children around 4 years of age, in contrast to 2- and 6-years-olds [4]. Finally, both active gestures (a child gestures itself) and passive gestures (gesturing performed by someone else) facilitate counting accuracy, although they lead children to make different types of errors [7].

Despite all that has been learnt about the supportive role of gesture in the acquisition of the counting skill, the questions about the exact nature of this contribution remain unanswered. A number of hypotheses concerning the issue have been brought forward, but based solely on behavioural data it is hard, if not impossible, to go much further. One of the tools that may be helpful in answering the questions *how* and *why* something happens (in addition to *what* happens),

is cognitive modelling. It turns out, however, that attempts to model the contribution of gesture to learning to count are virtually nonexistent. While a number of models can be quoted which, among other numerical capabilities, look at counting [10]–[12], their focus is mostly on the distinction between sequential enumeration and so-called subitizing (i.e. the immediate visual apprehension of small numerosities) and, more importantly, they do not address in any way the relation between counting and motor capabilities.

At this point it is worth to emphasise that the contribution of gesture to learning to count is an attractive topic from the point of view of embodied cognition [13], according to which, in general terms, the functionality of the brain cannot be understood without taking into account the body. There is a growing amount of evidence that, despite its abstract appeal, numerical thinking may be to a large extent shaped by physical interactions with the environment [14]. Understanding how the body contributes to the acquisition of such a concept as the number may well shed light on how abstract representations in general are constructed, an issue that is of vital importance for cognitive science [15]. It seems likely however that it is the embodied character of the contribution of gesture to learning to count that has been putting the researchers off from trying to model it. Indeed it is difficult to imagine a solely computational model which would not impose arbitrary assumptions about representing the bodily contribution. This

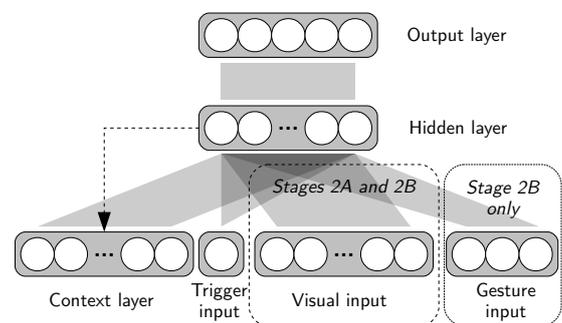


Figure 1. Architecture of the model. Gray polygons represent all-to-all connections. Activations are propagated from bottom to top.

general problem in the modelling of embodied phenomena is addressed in an elegant way by cognitive robotics [16], in which computational modelling is supplemented with an artificial, robotic body. This enables more accurate modelling with less arbitrary assumptions and the potential of this approach in the context of mathematical cognition modelling has already been demonstrated [17].

In this paper we attempt to fill-in the apparent gap mentioned above and make the first step toward understanding more specifically the contribution of gesture to learning to count. We propose a developmental robotic model of the phenomenon designed for the iCub humanoid robot platform [18], and investigate if the proprioceptive signal connected with gesture affects its counting accuracy. Furthermore we compare the behaviour of the model with data gathered in studies with human children.

The remaining of the paper is organised as follows. First, the model architecture, its development and evaluation procedures are described. Then a detailed statistical analysis of the experimental results is presented, including a comparison of the model behaviour with that of children. The paper concludes with a discussion mentioning perspectives for future work.

II. MODEL DESCRIPTION

In order to model the process of acquisition of the counting skill and the potential influence of the pointing gesture, we propose a recurrent neural network model based on the Elman architecture [19], presented in figure 1. In an Elman network, activations of the hidden units of a 3-layer artificial neural network at time $t - 1$ are available to the hidden units at time t (represented by the context layer), via connections which may be modified during training.

When formulating the model, the following assumptions about modelling the gesture have been made:

- proprioceptive information connected with gesture is an external input to the model. It is not the task of the model to *produce* gesture;
- gesture, if present, is a *correct* motor activity in the context of counting;

Motivations behind these assumptions are discussed in section IV. The following subsections describe the coding schemes adopted for model inputs and outputs.

A. Model task and output coding

The task used to assess the performance of the model throughout the experiments described in this paper is the production of a sequence of number words the length of which corresponds to the number of items presented to the visual input of the model. To this end, the output of the proposed artificial neural network is a binary vector of an a-priori chosen length (5 for the reported results), which encodes the produced words. Rather than adopting one-hot coding, output vectors are allowed to be non-orthogonal which is intended to mimic phonetic similarities present in natural languages. The particular sequence of number words to be used is a parameter of the modelling experiment and was generated randomly with

two constraints: different words need to have different vector representations and the special vector with zeros at every output unit corresponds to “silence” i.e. not producing any word. Real-life language data were not used, in order to enable the verification of stability of the model behaviour with respect to the sequence of number words. The model was trained and evaluated with numbers ranging from 1 to 10.

B. Input coding

1) *Trigger input*: The model has a 1-unit input called the trigger input. Its role is to indicate when the counting process should start. This corresponds to asking the subject in a psychological study a question (usually “How many?”) which according to the experimental protocol should encourage counting. Accordingly, the network is trained to remain silent irrespective of any additional inputs whenever the value of the trigger input is 0, and produce desired output when the trigger is 1. Trigger input activation remains fixed throughout every sequence in training and testing data sets.

2) *Visual input*: Visual input to the model is a 1-dimensional saliency map, which can be considered a simple model of a retina. Each unit of the visual input layer represents one spatial location in the visual frame of reference, and is activated or not depending on the presence of an object at this particular location. The sum of all activations over the visual input is normalised to 1 in order to eliminate the possibility of simple discrimination of cardinality based on this cue. Moreover, for a specific number of presented objects, the actual locations activated on the modelled retina are randomised between trials, so that the cardinality cannot be deduced based solely on locations of objects. The visual input consisted of 20 units to allow sufficient diversification of objects placement for the assumed maximum number (10).

3) *Gesture input*: The proprioceptive signal was obtained from a pointing gesture performed by the iCub humanoid robot [18]. As a starting point the data from joint angles of the robot right arm kinematic chain were used. This kinematic chain consisted of 6 degrees of freedom: torso yaw and pitch (roll angle was locked to eliminate unnaturally-looking postures) and the first 4 joints of the robot arm (shoulder and elbow). The robot was commanded, via its Cartesian interface, to point to 20 locations in front of it, which were assumed to correspond to the 20 locations in the visual input. These

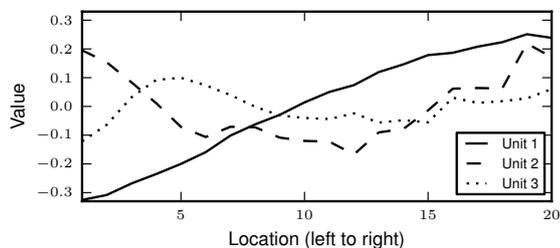


Figure 2. Values of the 3 units of the proprioceptive input (ordinate) for the 20 spatial locations (abscissa).

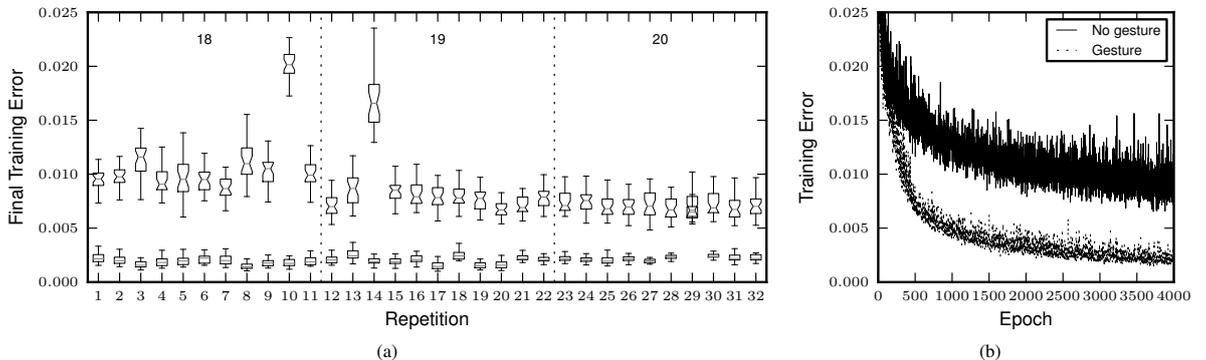


Figure 3. (a) box plot of the final training error (last 30 epochs) for each repetition in no gesture (notched boxes) and gesture (non-notched boxes) conditions. Outliers (data points outside the lower and upper quartile) were few and are omitted for clarity. Numbers near the top of the chart show the numbers of hidden units in trials; (b) typical example of full training error plots in both conditions for a selected trial (i.e. trial 1).

locations were uniformly distributed on a line placed 30cm in front of the robot, 10cm above its hip and spanned from 20cm left to 20cm right. Joint angles corresponding to pointing to each location were recorded and subsequently analysed using Principal Component Analysis which revealed that the first 3 principal components carry more than 90% of the total “statistical energy” of the original data. Therefore the dimensionality of the original signal was reduced from 6 to 3 by taking the 3 strongest principal components as the final proprioceptive input to the model. Values of these components for the considered 20 spatial locations are plotted in figure 2.

C. Model training

In order to model the development of the counting skill, the architecture of the artificial neural network (more specifically, its inputs) is adjusted throughout 3 modelling stages (see figure 1). First, the model is trained to recite a sequence of number words. Then, in order to assess the impact of the proprioceptive information connected with gesture, the training of the model branches out into two further stages, or conditions: the network is trained to count the number of objects shown to the visual input *in the absence* and *in the presence* of the proprioceptive gesture signal. All three stages use supervised learning by backpropagation through time and are described in detail below.

1) *Stage 1 – recitation of number words*: According to the findings of psychological studies with human children, early in the process of learning to count, children acquire a list of tags (i.e. number words) which is then used (more or less) consistently throughout the counting attempts [3]. In order to reflect this finding, the goal of the first stage of the model training was to teach it to produce a number word sequence, corresponding to numbers from 1 to 10. As illustrated in figure 1, at this point the only input to the model is the trigger input. The training data set consisted of two temporal sequences of the length of 20:

- the first sequence, with trigger input equal to 0 and all target outputs equal to 0 at every time step, trained the

model to remain silent when trigger input is 0;

- the second sequence, with trigger input equal to 1, trained the model to recite the 10 number words during the first 10 time steps of the sequence, and remain silent for the remaining 10 time steps;

The training lasted for 10000 epochs with learning rate 0.01 and weights updated in an on-line fashion.

2) *Stage 2A – learning to count without gesture*: In this training condition, the model is extended with the visual input, but the proprioceptive input is not added (figure 1). The task to be solved by the network is the same in both conditions (without and with gesture) and is to produce the sequence of number words with length equal to the number of objects shown to the visual input. Due to the previously mentioned need for randomisation of the positions of objects, the number of possible arrangements of objects grows exponentially with the visual input size, and thus it is highly impractical to create a data set containing all possible locations of objects for all considered numbers. In order to alleviate this problem, the network was trained in an “on-line” fashion, using smaller data sets which changed in every epoch and contained different arrangements of objects for a particular number. Each such small data set consisted of 22 sequences. For every number of objects on the retina ranging from 0 to 10 inclusive two sequences were included in the data set. For the first one, the trigger output was set to 0 and the target output consisted of 20 time steps of “silence”. For the second one, the trigger output was set to 1 and the target output contained the correct sequence of number words for the particular set of objects, followed by silence until the end of sequence. The arrangement of the objects was drawn randomly, but was the same for the two sequences in the small data set referring to a particular number.

3) *Stage 2B – learning to count with gesture*: Model training in the condition with gesture was analogous to stage 2A, main difference being the addition of the proprioceptive gesture input (figure 1). The training was also performed using

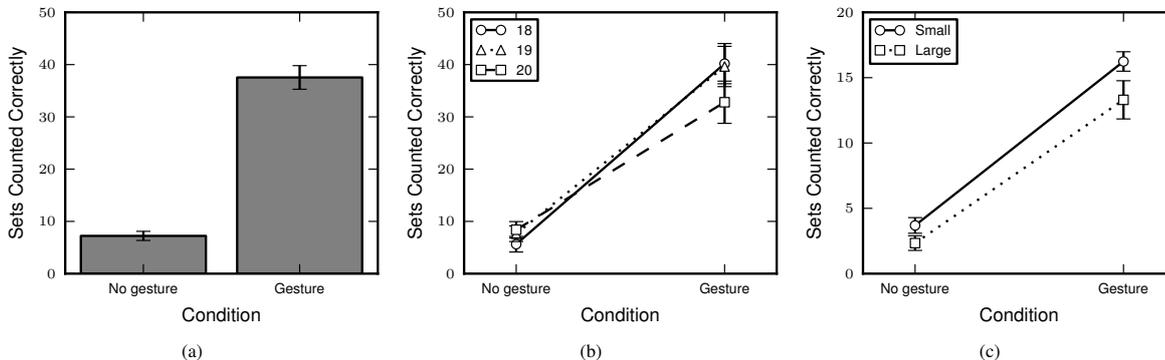


Figure 4. (a) mean number of collections from the evaluation data set counted correctly (out of 50) by the model without and with gesture; (b) profile plot of the mean number of collections counted correctly (out of 50) without and with gesture for models with 18, 19 and 20 hidden units. Visible interaction reached statistical significance (see main text); (c) profile plot of the mean number of collections counted correctly (out of 20) for small (1-4 items) and large (7-10 items) collections, without and with gesture. On all charts error bars indicate ± 2 SEM.

small datasets, however the sequences in the dataset contained an additional gesture signal constructed as follows. For the given arrangement of objects in the visual input, the locations of objects were considered in the left-to-right order. Assuming there were n objects on the retina, for time steps $t < n$, the proprioceptive input consisted of the joint angles (reduced to 3 dimensions as explained above) corresponding to the position of the retina at which the $(t + 1)$ -th object (in the left-to-right order) was located. For $n \leq t < 20$, proprioceptive input remained unchanged with respect to the values present for the last object. For all sequences for which the trigger input was set to 0 or the number of elements on the retina was 0, no gestural signal was provided (all 3 proprioceptive inputs set to 0). In stages 2A and 2B, training lasted for 4000 epochs using the same learning rate as stage 1.

D. Model evaluation

Evaluation of the model performance has been designed with an intention to yield maximum similarity to studies with human participants (more specifically, to [7], where authors investigate experimentally the function of gesture in learning to count with special focus on the distinction between keeping track and coordination of the recited number words with counted items), in order to allow subsequent comparison of the behavioural data. To that end, the 5-dimensional output of the model was first transformed into number words by means of nearest neighbour classification. Then, the resulting number words sequence was compared in detail with the target sequence (corresponding to the correct counting of presented items), using the same criteria of correctness and categorisation of possible errors as applied by [7]. Thus, the following counting errors were distinguished (see table 1 in [7]):

- Skip: the model does not assign a number word to an object;
- Continue: the model continues to count beyond the number of objects shown;

- Stop short: the model does not assign a number word to the last object(s);
- String error: the sequence of produced number words contains any other error;

Because of the nature of the model and the assumptions made about the gesture signal, “Double count” and “Distracted” errors described by [7] are not applicable in this study.

The data set used for model evaluation was independent from the training data sets and contained 50 sequences (5 for every number ranging from 1 to 10) with randomised locations of objects. The same evaluation data set was used to test the model in both gesture and no gesture conditions. However, consistent with the assumptions about the proprioceptive input, when the model trained in stage 2B was evaluated, sequences in the evaluation data set included the corresponding proprioceptive gesture signal.

In order to answer the question whether the addition of gesture to the designed model during training improves final counting accuracy, 32 independent repetitions of the model training were performed. In all trials the same sequence of number words was used, which is intended to correspond to the situation of testing children who use the same language. In order to verify the robustness of results with respect to the number of hidden units in the model, 18, 19 and 20 hidden units were used in 11, 11 and 10 of the trials respectively.

III. EXPERIMENTAL RESULTS

As mentioned above, the experimental set-up in this study was intentionally designed to resemble the one applied by [7]. Therefore, the analysis of the results of experiments with the proposed model presented below resembles in many aspects the one found in the quoted paper.

A. Learning the sequence of number words

In all 32 trials, the model successfully achieved the prerequisite task, i.e. has learnt to remain silent when the trigger input is 0, and to produce the correct sequence of number

words when the trigger input is 1. The final backpropagation through time training error (i.e. the mean squared error over all sequences in the training set) obtained in all trials was low, with the maximum $1.27 \cdot 10^{-4}$ (for repetition 14). In other words, in all trials the model learnt to perform the prerequisite task very well, despite differences in numbers of hidden units.

B. Gesturing and counting accuracy

Investigation of the effect of the proprioceptive gestural input on counting accuracy starts with a look at the progress of the model training in both no gesture and gesture conditions. In figure 3a a box plot of the training error for the last 30 epochs in both conditions for each trial is shown. For all but 1 repetition (29), the final training error in the gesture condition was considerably lower than in training without gesture. A typical plot of the training errors throughout the full stage 2A/B is shown in figure 3b.

Similarly to [7], we present an analysis of the correctness of the counting sequences produced by the model. In figure 4a the mean number of sets of objects from the evaluation data set counted correctly by the model in no gesture and gesture conditions is shown (this is analogous to Figure 1 in [7]). Statistical analysis of the results was performed in the form of a 3×2 (18, 19 and 20 hidden units times no gesturing and gesturing) repeated measures MANOVA with the gesturing condition as the within-subject factor (since for every experiment repetition the same model originating from the stage 1 of training was used in stage 2A and 2B) and the number of hidden units as the between-subject factor (in order to assess the stability of results with respect to the number of hidden units in the model). Comparison of the mean number of sets counted correctly for condition without and with gesture was a planned within-subject contrast. The analysis indicated strong statistical significance of the difference ($F = 633.686$, $p < 0.001$) between the gesturing conditions, meaning that models trained with proprioceptive gesture input available counted more collections in the evaluation data set correctly than those without this input. This is in perfect agreement with findings reported by [7], where children’s performance in the no gesture condition was significantly inferior to conditions with gesture. The between-subject effect of the number of hidden units in the model was not found ($F = 2.194$, $p = 0.13$), indicating that the beneficial effect of gesture was robust within the considered range of numbers of hidden units. However it has to be acknowledged that the effect of within-subject interaction between gesturing condition and the number of hidden units approached statistical significance ($F = 6.228$, $p = 0.006$). This means that in the conducted experiments the beneficial effect of the additional proprioceptive input was stronger for neural networks with less hidden units. This conclusion is illustrated in the profile plot of the estimated marginal means for gesture condition versus the number of hidden units shown in figure 4b.

Next step of the analysis presented in [7] focused on the dependence of the effect of gesturing on the size of the counted collection. In order to investigate this in the proposed model,

Table I
MEAN NUMBER OF CORRECTLY COUNTED COLLECTIONS (OUT OF 20)

Condition	Small sets	Large sets
no gesture (2A)	3.691 (0.299)	2.339 (0.280)
gesture (2B)	16.239 (0.372)	13.303 (0.731)

Table II
MODEL COUNTING ERRORS

	% of trials (out of 1600) with error made		% of models (out of 32) which made error	
	no gesture	gesture	no gesture	gesture
Skip	0.8	0.2	37.5	37.5
Continue	43.4	7.1	100.0	100.0
Stop short	39.3	14.4	100.0	100.0
String	13.1	3.9	93.8	93.8

the collections from the evaluation data set were divided into small numbers (1-4) and large numbers (7-10) and statistical analysis of the mean number of sets counted correctly within these groups was performed (meaning this time $3 \times 2 \times 2$ MANOVA). Obtained values are summarised in table I (which corresponds to Table 2 in [7]). In this table, numbers in parentheses show the standard error. Once again the behaviour of the model is in perfect agreement with the psychological study. Strong effects of gesture and set size were found ($F = 597.736$ and $F = 31.814$, respectively, $p < 0.001$ in both cases), while there was no interaction between these two factors ($F = 2.905$, $p = 0.099$). This means that the proposed model, similarly to children, counts small sets more accurately than large sets, and that gesturing improves its counting accuracy for both small and large collections of objects. This is illustrated in a profile plot in figure 4c.

C. Error patterns

Finally we look more closely at errors made by the model when counting without and with gesture. Table II reports the percent of trials with particular types of errors as well as the percentage of models which made particular kinds of errors. This is similar, but not equivalent to Table 3 published in [7], which focuses on the differences between active and passive gesture while herein such distinction is not made. When considering the general picture however, one can conclude with a fair degree of certainty that overall, patterns of errors made by the proposed model are different from the ones obtained in the study with children. While for children the most common errors are Skip and Double count, the model proposed in this paper makes Continue and Stop short errors more often. In contrast to 5% and 15% of children tested by [7] who committed a Continue and Stop short error at least once, the model made these errors at least once in every trial.

IV. CONCLUSIONS AND DISCUSSION

A recurrent artificial neural network model of the contribution of gesture to learning to count has been proposed. In an experimental set-up designed to allow comparison with human

data it has been confirmed that the proprioceptive gesturing signal enabled the model to improve its counting accuracy. The model behaviour yielded similarity to that of human children in terms of the effects of gesture and of the counted set size, although the obtained patterns of errors were different.

A few issues about the proposed model are discussed below. The first are the assumptions regarding the gesture signal (section II). In this study the gesture is an external input to the model. Although it may at first seem arbitrary and artificial, such approach is in line with a finding that children, when counting, apply the one-one correspondence principle in gesture before it is transferred to speech [1]. While designing a model which produces gesture is planned for future work, the focus of the present study was to test, from an “information-theoretic” point of view, if the proprioceptive signal carries information which may be exploited when learning to count. Obtained results confirm this is indeed the case.

Second, the proposed model makes different counting errors than human children. This may have been caused by two properties of the chosen model architecture. The assumption that the gesture is always correct affects the kinds of errors that may appear. More specifically, Double count errors do not appear at all, and Skip errors are also affected (although not ruled out). This may be seen as corresponding to the “puppet condition” in the study [7] where gesture performed by the puppet was also always correct. In addition, error patterns produced by the model may be influenced by time discretisation which is an inherent property for the Elman architecture. Here, the synchrony between gesture and number words recitation is naturally present. However, according to some hypotheses, synchronising the number words production with tagging the objects being counted may be one of the major functions of gesture [6]. Therefore, a model with continuous time would be more appropriate to investigate the importance of synchrony in the context of counting, and this is also included in the plans for future work. Error patterns would likely change as 4 out of 5 considered error types may appear as a result of problems with synchronisation.

Finally, the scalability and generalisability of the obtained results need to be addressed. As mentioned before, performed 32 experiment repetitions used the same sequence of number words. It has been confirmed however in earlier informal experiments with the model that the reported results are not due to any specific characteristic of the used number words sequence. These tests were also used to establish the training parameters (e.g. the number of epochs and the number of hidden units) for which the training of the model on the target task is successful. The crucial findings of this paper, i.e. the effect of gesture and set size on the counting accuracy, should hold for any reasonably chosen number words sequence length or retina size, as the particular values of these parameters were chosen arbitrarily and do not lead to any loss of generality. Of course this holds provided that the model architecture (most importantly the number of hidden units) and the training parameters are also adjusted accordingly.

Present study provides quantitative evidence in support of

the intuition that motor knowledge connected with pointing gesture can be transferred to a verbal and conceptual competence. Cognitive robotics approach to modelling alleviates the need for arbitrary assumptions about representing the proprioceptive signal, since a real robot which performs the actual pointing gesture is available. The analysis of the internal workings of the model and employing more sophisticated models are expected to shed even more light on the nature of the contribution of gesture to learning to count.

ACKNOWLEDGEMENT

This research has been supported by the EU project RobotDoC (235065) from the FP7 Marie Curie Actions ITN.

REFERENCES

- [1] T. A. Graham, “The role of gesture in children’s learning to count,” *Journal of Experimental Child Psychology*, vol. 74, no. 4, pp. 333–355, 1999.
- [2] B. Schaeffer, V. H. Eggleston, and J. L. Scott, “Number development in young children,” *Cognitive Psychology*, vol. 6, no. 3, pp. 357–379, 1974.
- [3] R. Gelman, “What young children know about numbers,” *Educational Psychologist*, vol. 15, no. 1, pp. 54–68, 1980.
- [4] G. B. Saxe and R. Kaplan, “Gesture in early counting: A developmental analysis,” *Perceptual and Motor Skills*, vol. 53, no. 3, pp. 851–854, 1981.
- [5] R. Gelman and E. Meck, “Preschoolers’ counting: Principles before skill,” *Cognition*, vol. 13, no. 3, pp. 343–359, 1983.
- [6] K. C. Fuson, *Children’s counting and concepts of number*, ser. Springer series in cognitive development. New York, NY, US: Springer-Verlag Publishing, 1988.
- [7] M. W. Alibali and A. A. DiRusso, “The function of gesture in learning to count: More than keeping track,” *Cognitive Development*, vol. 14, no. 1, pp. 37–56, 1999.
- [8] R. A. Carlson, M. N. Avraamides, M. Cary, and S. Strasberg, “What do the hands externalize in simple arithmetic?” *Journal of Experimental Psychology-Learning Memory and Cognition*, vol. 33, no. 4, pp. 747–756, 2007.
- [9] J. Piaget, *The child’s conception of number*. Oxford, England: W. W. Norton & Co., 1952.
- [10] P. Rodriguez, J. Wiles, and J. L. Elman, “A recurrent neural network that learns to count,” *Connection Science*, vol. 11, no. 1, pp. 5–40, 1999.
- [11] S. A. Peterson and T. J. Simon, “Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture,” *Cognitive Science*, vol. 24, no. 1, pp. 93–122, 2000.
- [12] K. Ahmad, M. Casey, and T. Bale, “Connectionist simulation of quantification skills,” *Connection Science*, vol. 14, no. 3, pp. 165–201, 2002.
- [13] R. Pfeifer, M. Lungarella, and F. Iida, “Self-organization, embodiment, and biologically inspired robotics,” *Science*, vol. 318, no. 5853, pp. 1088–1093, 2007.
- [14] G. Lakoff and R. Núñez, *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books, 2000.
- [15] L. Barsalou, “Perceptual symbol systems,” *Behavioral and Brain Sciences*, vol. 22, no. 04, pp. 577–660, 1999.
- [16] M. Asada, K. MacDorman, H. Ishiguro, and Y. Kuniyoshi, “Cognitive developmental robotics as a new paradigm for the design of humanoid robots,” *Robotics and Autonomous Systems*, vol. 37, no. 2–3, pp. 185–193, 2001.
- [17] M. Ruciński, A. Cangelosi, and T. Belpaeme, “An embodied developmental robotic model of interactions between numbers and space,” in *33rd Annual Meeting of the Cognitive Science Society*, L. Carlson, C. Hoelscher, and T. F. Shipley, Eds., 2011, pp. 237–242.
- [18] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, “The iCub humanoid robot: An open-systems platform for research in cognitive development,” *Neural Networks*, vol. 23, no. 8–9, pp. 1125–1134, 2010.
- [19] J. Elman, “Finding structure in time,” *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.