Evolutionary Robotics Simulation Models in the Study of Human Behaviour and Cognition

Marieke Rohde

Submitted for the degree of D.Phil.

University of Sussex

April, 2008

Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Signature:

Acknowledgements

Most importantly, I want to thank my supervisor Ezequiel Di Paolo who has supported me on so many levels - this thesis would not be what it is without him. He dynamically switched between the roles of my academic supervisor - passionately engaging with my research as if it was his own; providing technical, conceptual, formal and strategic support - my mentor - training me for the academic profession; involving me in reviewing, organisation of academic events; promoting me as his substitute for invited lectures - my PA, compensating for my scatter-brainedness; reminding me of deadlines and duties, spell-checking and smoothing out anything I wrote - and my friend; supporting me through emotional and economic crises; bearing with me during phases of egocentricity and self-obsession; sharing his thoughts and ideas with me, serious and silly, over a cup of coffee or a pint - or five.

Thanks to CCNR for discussions, moral support and general helpfulness. Special thanks go to the Bramber house regulars (in variable cast over the years), for making me laugh so much, for table football and for coffee in the sun. The fellow geek-girls for coffees, walks, being great friends and always there when I needed them. Andy - I owe you one, or maybe several.

I thank the GSP for being my academic home during the five months in Compiègne. Olivier, Charles and John for inspiration, guidance, support, friendship and fantastic food. Dominique for turning my ideas into reality. Amal, Dominique, Bart and Yusr for instructing me in French humour and language and for making sure that I stay human during lunch and coffee breaks.

Additional to the above, I want to thank all the passionate and intelligent individuals, teachers and peers, who shaped my mind, intellect and personality, in school, in the SchülerAkademie, in the Leibniz Kolleg, in the IKW of Osnabrück University, in the Cognitive Science degree in the University of Costa Rica, in the InDy group of the Fraunhofer AIS, on the EASy MSc at Sussex and, most importantly, outside academic context.

I want to thank those who have emergency reviewed parts of this dissertation (in priority order): Ezequiel, Hanne, Mike B., Greg, Benoît, Pete, Matthew E., Simon, James, Tom, Eduardo, José and Al.

A huge thank you goes to my housemates who have borne with me over the last 3.5 years. Uschi, not only for our incredibly smoothly functioning flat-share, which made home a true place to recover and recharge the batteries, but also for being an intelligent and common sense commentator on my ideas and work (and life). Uschi, Catherine and Philine for friendship, support and open ears, for wine, silliness and for our by times quite ridiculously intellectual leisure activities.

My family - particularly my mum. Where would I be had she not fed me, sorted through stuff, emergency sent me documents or taken control when I was hysterical as if I was still a little girl?

Finally, I want to thank my friends near and far. In particular, I want to thank Michelle Beltran and Crisi Kabisch because they have helped me through the most difficult times during these last years. For the others - I do not need to name you. You know who you are, how much you mean to me, and how much I need and appreciate you!

Evolutionary Robotics Simulation Models in the Study of Human Behaviour and Cognition

Marieke Rohde

Summary

Simulation models are powerful scientific tools for embodied, behaviour-based, dynamical or enactive approaches in Cognitive Science that reject the traditional metaphor of the mind as a digital computer. However, simulation models in general and Evolutionary Robotics simulation models in particular are frequently deliberately minimal. If they are used in the study of human level cognition, there appears to be a gap between the simplicity of the model and the sophistication of the *explanandum*. This thesis presents a number of models developed to elucidate different problems in Cognitive Science and which exemplify the different scientific values that Evolutionary Robotics simulation models can have for the study of human intelligence and behaviour (i.e., hypothesis generation, proof of concept, verification and illustration beyond cognitive limits and intuitions).

Applying this simulation method to problems in motor control, neuroscientific theory, social contingency and time perception, it is argued that, in order to take part in explaining human behaviour and cognition, it is not necessary - or even desirable - to aspire a model that approximates human level complexity. Besides the results each of the models presented contributes to the research question it addresses, the main result from the work presented in this dissertation is the proposal of a new and interdisciplinary methodological framework. This framework combines Evolutionary Robotics simulation models and experiments in human sensorimotor adaptation, using the same minimal computer simulation in both the experiment and the model. The application of this method to the problem of adaptation to sensory delays and perceived simultaneity demonstrates its potential to explain the sensorimotor basis of perception in a *bona fide* enactive way.

Submitted for the degree of D.Phil.

University of Sussex

April, 2008

Contents

List of Abbreviations and Symbols

α_i	Joint angle of i^{th} joint
a_i	Activation of unit n_i
AI	Artificial Intelligence
ALife	Artificial Life
ANN	Artificial Neural Network
BBR	Behavior-Based Robotics
C, c_{ij}	Network connectivity matrix in which $c_{ij} \in \{0, 1\}$ indicates existence of a connection from unit n_j to unit n_i
CCNR	Centre for Computational Neuroscience and Robotics, University of Sussex (host institution)
CPG	Central Pattern Generator
CTRNN	Continuous-Time Recurrent Neural Network
δ, Δ	Parameters of RBF (see chapter 4)
d	Delay (of sensory inputs to CTRNN controller)
$d(x)$	A distance function (locally defined)
DS, DST	Dynamical System, Dynamical System Theory
DoF	Degree-of-Freedom
ε	Noise or a very small constant (locally defined)
ER	Evolutionary Robotics
ϕ	Required pointing direction signal (see chapter 4)
$F(i)$	Fitness function for individual i, performance for participant i
FLE	Flash-Lag-Effect in psychophysics
GA	Genetic Algorithm
GOFAI	Good-Old-Fashioned Artificial Intelligence
GSP	Groupe Suppléance Perceptive, Université de Technologie de Compiègne
h	Simulation time step
I_i	External input to n_i
k_i	A constant
$K(\phi)$	Linear synergy function (see chapter 4)
M_i	Motor signal
M_G	Motor gain
MSE	Mean Square Error
n_i	The i^{th} unit (neuron) in an ANN/CTRNN
ω	Angular velocity
ODE	Open Dynamics Engine (C++ library)
PDP	Parallel Distributed Processing

PS	Perceptual Supplementation
r	Magnitude of vector mutation in GA
RBF,RBFN	Radial Basis Function, Radial Basis Function Network
σ	Standard deviation
$\sigma(a)$	Standard logistic (sigmoidal) function
S_i	Sensory signal
S_G	Sensor gain
θ_i	Bias of unit n_i
τ_i	The time constant of decay of a_i
t, T, t_0	t = time, T = length of task, t_0 = initial/reference time
TLP	Wittgenstein's *Tractatus Logico Philosophicus*
TM	Turing Machine
TNGS	Theory of Neuronal Group Selection
TVSS	Tactile Visual Sensory Substitution
v	velocity
W, w_{ij}	Network weight matrix in which w_{ij} gives the connection weight from unit n_j to unit n_i
x^*	Fixed point or steady state activity (numerically established) of variable x in a DS

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation essentially promotes *enaction* as an alternative paradigm for Cognitive Science and Evolutionary Robotics simulation models as one of its tools. The main objective is of a methodological nature i.e., to provide answers to the question: *how can minimalist Evolutionary Robotics simulations be used to explain human (high-level) cognition?*

Before explaining how this question is posed, addressed and answered, however, I want to start with an autobiographic account, explaining how I came into the position from where I started the research journey documented here. This journey took me from modelling motor control of the arm, via philosophising about the plausibility of neural architecture to the interdisciplinary study of perceptual experience combining synthetic robotics modelling with minimal experimental work on human sensorimotor behaviour. Alongside the corpus of concrete results from the hands-on modelling and experimental work, the methodology applied kept growing, developing and improving and, at the end of the journey, I realise that this leitmotif is possibly its main result. The personal account in section 1.1 presumes some familiarity with the paradigmatic debate in Cognitive Science, which is recapitulated for the naïve reader in chapter 2. Section 1.2 presents the structure of the current document.

1.1 Genesis

Many of my colleagues and friends know that I have studied Cognitive Science for three years, completely oblivious about the existence of alternative paradigms in Cognitive Science. I quite happily learned and taught LISP and Prolog, implemented Montague semantics for 'formal semantics' and hidden Markov models for 'natural language processing'. I learned brain areas by heart for 'Functional Neuroanatomy' and drew boxes in 'Cognitive Psychology'. I wrote essays about the philosophy of mental representations and analysed the advantages of concept lattices in knowledge representation. The only thing I was not happy about was that two and a half year into my degree, I had learned very little about how the mind works.

This is, obviously, an ironic exaggeration of my experience during the years of my undergraduate studies. These years were probably the most inspiring, educatory and formative of my life. I learned so much about the brain, about programming, logics, computability and useful AI tech-

niques, about structural properties of language, about statistics, philosophy, experimental design and psychology. All in all, it was a stimulating atmosphere in which I was privileged to study. The problem was more that things never seemed to *come together*. I am proud to say that I have first person experience of frustration with the computationalist paradigm in Cognitive Science.

What most people do not know is that the seeds for what happened next had already been planted much earlier than that. I struggle to say when exactly, but twelfth grade philosophy class was certainly a milestone. Kant's transcendental epistemology mesmerised me. The idea of the inaccessible *Ding an sich* (thing in itself) that is so difficult to grasp seemed to stab right into the heart of some diffuse doubts I had held about myself and the world and that had been silently bugging me. It was the same time that we learned about relativity theory and Einstein's mind boggling thought experiments in physics class. When we were given an excerpt by von Glasersfeld[1] about the construction of time and space as the distinction of the concept of 'after' from the concept of 'next to', based on our sensorimotor experience, I got even more excited. This was like Kant, but better. Going all the way.

The problem at the time was that nobody seemed to share my enthusiasm. I kept my fancy for radical constructivism alive despite dismissal by peers and teachers till into the first year of my studies. My first ever term paper (during College in Tübingen) was on time and space in Kant and von Glasersfeld and how it relates to Newtonian as opposed to modern physics. My second term paper (in Cognitive Science: 'philosophy of language') was a critique of Davidson's refutation of solipsism and an appraisal of radical constructivism. Just as my philosophy teacher, the lecturers marking these essays showed themselves more than impressed with my knowledge, understanding and reasoning but had little sympathy for my conclusions. I got bored and gave up on constructivism - temporarily.

It took a bit of luck for me to get back on course. At the end of a rather software-engineering oriented two months internship at the Fraunhofer Institute for Autonomous Intelligent Systems, I strolled, out of curiosity, over to the robotics section of the building and started chatting to people about my studies, my interests. I ended up talking to Pasemann and colleagues at the Intelligent Dynamical Systems unit and left with a contract to come back for my BSc project. Another milestone. During those months, I learned to work with Evolutionary Robotics simulations and Dynamical Systems analyses. Reading Pasemann's *Repräsentation ohne Repräsentation* (1996), I realised that, through the failure of AI, constructivism was gaining ground.

The rest is boring. It was the logical next step to come to Sussex, the cradle of Evolutionary Robotics and Artificial Life, for the MSc in Evolutionary and Adaptive Systems. I learned about autopoiesis, enaction, Artificial Life, GAIA theory, sensorimotor plasticity etc. I mastered the crisis suffered annually by students on the course once you realise there are no walking talking robots and I learned to recognise the merits of minimal Evolutionary Robotics simulation modelling, the elegance with which these simple models proved that representationalism was wrong, even if it is yet another phototactic agent, and was offered a place to do a DPhil.

Still, I was diffusely dissatisfied. Sure, it was not our aim to evolve a human. Sure, it feels good to be right. But what I really wanted to know is how the mind works. Any answers to this question appeared to come from hands-on scientific work, Kohler's (1962) experiments with

[1] I have not found this source again till the present day.

goggles, Held's (1965) work on sensorimotor plasticity, Bach-y-Rita's (2003) research on Sensory Substitution, Piaget's (1936) classical work on conceptual development, Núñez' (e.g., Núñez & Sweetser, 2006) anthropological studies. What did Evolutionary Robotics contribute, or the area of Artificial Life in general? Was 'proving wrong' the only thing our models could do? Was I doing the wrong thing?

There were examples of Evolutionary Robotics models that made concrete and constructive contributions to a field of research: for instance, in explaining insect behaviour (e.g., Vickerstaff & Di Paolo, 2005), in developmental psychology (e.g., Wood & Di Paolo, 2007) or in motor control (e.g., Rohde & Di Paolo, 2005, also chapter 4 of this dissertation). But, even though ultimately related, these contributions seemed very remote from the kinds of questions I was really interested in. Even if the computationalist Cognitive Science was misguided, could it be that our critics from the traditional camp were empirically right in claiming that ALife modelling was a dead end?

Posing and refuting this question is the main purpose of the current dissertation. It was the guiding principle that has led my research over the past three and a half years. Five Evolutionary Robotics simulation models are presented here, all of which address questions of human cognition and behaviour and provide valuable - and valu*ed* - results. The model of motor synergies strongly idealises experimental work on human grasping behaviour in order to prove that the postulated motor control principles can work even in abstracted and modified conditions and to generate hypotheses for further experimental work. The model of value system architectures illustrates conceptual weaknesses in certain types of neuro-cognitive architecture and points out implicitly held prior assumptions that need to be explicated. The two models of human perceptual crossing in minimal simulated environments generate proofs of concept and hypotheses on different levels of explanation that feed back into experimental practice and theory building. This modelling of perceptual crossing led up to the interdisciplinary project on adaptation to sensory delays, in which the model plays the same methodological role as before, however, it does not model experimental work by other researchers but was developed alongside my own experimental study, which I conducted during a five months research stay at the *Groupe Suppléance Perceptive* in Compiègne. This group had also conducted the mentioned experiments on perceptual crossing. With this body of results from isolated modelling activity as well as from modelling activity as part of an interdisciplinary methodology, I am happy to conclude that I have been doing the right thing: Evolutionary Robotics modelling is a great tool for the study of human cognition and behaviour.

The current dissertation is the document of this crooked road with motivational ups and downs, that passes through areas of technical, experimental, social, linguistic and analytic activity, detouring left and right, touching on seemingly unrelated questions across disciplines. Having arrived at the end of this road, I am surprised to find myself right were I started: contemplating how time and space are constructed from sensorimotor experience. The difference is, however, that I have a whole trunk full of knowledge and ideas on how to advance with this question, which include Evolutionary Robotics simulation modelling as an essential and valuable tool to investigate cognition and perception in a *bona fide* enactive way - no reductionist charlatanry.

1.2 Structure of the Current Dissertation

This section briefly summarises the contents and structure of the remaining twelve chapters and how they work together. The dissertation basically consists of two parts. Chapter 2-7 present the ideological and methodological background, develop the novel methods proposed and present results from Evolutionary Robotics simulation models that address different kinds of questions across disciplines. The second part (chapters 8-12), in contrast, focuses on one particular research question, i.e., the study of adaptation to sensory delays and perceived simultaneity. This second part applies the interdisciplinary approach that results from the first part at a larger scale, combining experimental and modelling results. The conclusion chapter with the unlucky number 13 comes back to the methodological theme of the dissertation and evaluates the results from all experimental and modelling chapters in the light of the conceptual debate outlined in chapter 2 and chapter 3.

The following chapter (chapter 2) is an introduction to the paradigmatic struggle in Cognitive Science. It starts off by giving a historical account of the birth, rise and fall of computationalist Cognitive Science. It then introduces some of the main proposed alternatives and clarifies how they differ from each other and from the computationalist paradigm. Then, the enactive paradigm is outlined and advocated in more detail. The research question of this dissertation, i.e., how Evolutionary Robotics modelling can contribute to an enactive explanation of mind, is presented in the context of criticism and challenges to the enactive approach in general and concerning the role of Evolutionary Robotics simulation modelling in particular.

Given that this dissertation is, in the first place, a methodological dissertation, it is not surprising that the method chapter 3 is the longest chapter. It does not only introduce the technical details of the experimental and modelling approaches taken. It also presents original contributions on a number of science theoretic questions, such as the consequences of a constructivist world view for scientific practice and interpretation, the role of Dynamical Systems Theory in the enactive approach and the methodological difficulties associated with the scientific study of experience. It concludes with the outline of how minimal experimental (Perceptual Supplementation) and modelling (Evolutionary Robotics) approaches can be integrated to form a minimal and interdisciplinary method to address questions of perceptual experience from the enactive perspective.

Chapter 4 presents a simulation model studying linear synergy as a principle in motor control. The problem of redundant degrees of freedoms and the concept of motor synergies are introduced, as well as the experimental study that inspired the simulation model. After presenting the model and the results, the study is evaluated both as to what it implies for the research question it addresses as a self-contained model and in the light of the research question of the current dissertation, i.e., if and how Evolutionary Robotics simulation models can contribute to the study of human cognition and behaviour.

A simulation model caricaturing value system architectures in order to illustrate the implicit premises underlying this kind of architecture is at the core of chapter 5. However, the introduction section of the chapter does not only provide the necessary background for posing the question addressed by the model but also outlines some novel contributions about value generation and the origins of meaning in the enactive approach that are independent of this model. Again, the model and its results are presented and consequently evaluated with respect to the question the model

addresses as well as with respect to the methodological theme of the dissertation.

Chapters 6 and 7 present the results from two simulation models of two subsequent and very related experimental studies on human perceptual crossing in a one-dimensional (chapter 6) and a two-dimensional (chapter 7) minimal virtual environment. Even though I was not personally involved in the experimental work underlying these simulation models, these chapters already implement the proposed combination of Evolutionary Robotics modelling and minimal Perceptual Supplementation experiments that I propose in chapter 3. The models and their results are evaluated, identifying their contribution to explaining the dynamics of perceptual crossing and in the light of the overarching methodological theme.

The purely conceptual chapter 8 can be seen as a second introduction chapter, framing the question addressed in the second part of the dissertation. It summarises and relates work on time cognition and construction from a broad variety of sources, including Kant's epistemology, Husserl's and Merleau-Ponty's phenomenology, Lakoff and Núñez anthropology, Piaget's developmental psychology, Varela's neurophenomenology, Shanon's study of altered states of consciousness, Libet's neuroscientific work on neuro-behavioural latencies and Nijhawan's psychophysics study of the flash lag phenomenon. Again, the chapter does not only review but presents novel thoughts and ideas. It then presents, more specifically, related work on adaptation to sensory delays, in particular the psychological study that inspired the second part, in order to phrase the hypothesis tested in the experiment.

In chapter 9, the experimental results from the study on human adaptation to tactile delays in a minimal virtual environment are presented. The results do not confirm the initial main hypothesis. A preliminary data analysis is performed, leading to ambiguous results that suggest non-trivial reasons for the failure of the main hypothesis, which deserve further exploration.

The Evolutionary Robotics simulation model of the experiment is presented in chapter 10. The results clarify the dynamics of the task posed to the experimental participants, and the impact of strategy on patterns of adaptation. There are quantifiable differences in behaviour between agents that solve the task with and without delays that relate to dynamical principles governing the task. These differences suggest the investigation of the same variables in the experimental data gathered.

In the light of these simulation results, the sensorimotor recordings from the experiment are revisited in chapter 11. Looking at the variables pointed out to be potentially relevant by the simulation model, it can be shown that the behaviour of the experimental participants follows the same regularities along some dimensions but not along others.

Chapter 12 evaluates the experimental and simulation results and sketches how, in the light of the conceptual analysis in chapter 8, the theoretical and empirical insights about the sensorimotor principles of the task lead to new hypotheses for further experimentation. New questions are phrased and old ones rephrased, providing the basis for a future research program on the study of simultaneity perception and how it is influenced by adaptation to sensory delays.

The conclusion chapter 13 returns to the main theme and assesses the methods developed and employed and in how far they can and do contribute to the study of human cognition and behaviour. Even though the conclusion is predominantly positive and optimistic, weaknesses and limits of the approach taken are self-critically assessed and presented.

Chapter 2

An Enactive Approach to Human Cognition and Sensorimotor Behaviour

---

This opening chapter introduces the philosophical and paradigmatic context in which the current dissertation and the research it presents have been generated. It forms the foundation for the description and development of the methods employed and developed (chapter 3) and their later application (Just modelling: chapters 4 - 7. Combined modelling and experimental work: chapters 8-12). The significance of the results of each of the models and experiments for the particular research question they address is discussed within the respective experimental or modelling parts of the dissertation. The paradigmatic and methodological implications of these studies, which are the unifying research theme for the present work are identified and evaluated in the conclusion 13.

In many ways, the methodological research question of this dissertation can be seen as yet another episode of the paradigmatic struggle between traditional computationalist Cognitive Science and more embodied and dynamic approaches. Therefore, this chapter starts (section 2.1) with a summary of the key issues, persons and milestones that have determined this debate, which is as old as Cognitive Science itself. In Cognitive Science, there is a tendency to present the paradigm struggle as a battle between the traditional 'GOFAI' (good-old-fashioned Artificial Intelligence (Haugeland, 1985)) approach on the one hand and everything which is '$\neg$ GOFAI' (or 'New AI') on the other hand. Even though alternative proposals (Connectionism, Dynamicism, Behaviour-Based Robotics, ...) have frequently originated from the same observations of shortcomings of the traditional paradigm and often have significant methodological and ideological overlap, they cannot be seen as a single alternative that comes in different flavours. Significant tensions exist between them. Section 2.2 summarises a number of alternative paradigms, identifies their maxims and core assumptions and points out in how far they are prone to the same criticisms as GOFAI. Section 2.3 presents the enactive approach as the paradigm in Cognitive Science that this dissertation is based on. Finally section 2.4 is a self-critical reflection on the main challenges this paradigm faces and particularly on the role computational models have played in it. Special attention is paid to a criticism that dynamical modelling approaches frequently face, i.e., that such models serve well to address low-level behavioural issues but not high-level cognitive issues. This last section finishes by outlining the challenge that has driven the research conducted in this dissertation, i.e.,

to identify ways to use simple Evolutionary Robotics simulation models in Cognitive Science in general and, in particular, for the scientific study of high-level cognition.

2.1 The Rise and Fall of Traditional Cognitive Science

To my knowledge, it is not clear when the term 'Cognitive Science' was first employed. Its birth is, however, frequently associated with the birth of a more traceable term, i.e., 'Artificial Intelligence' (AI; e.g., Eysenck & Keane, 2000; Haugeland, 1981; Russell & Norvig, 1995), a label that has first been used for in the call for the Dartmouth Conference in 1956 (McCarthy, M.Minsky, Rochester, & Shannon, 1955). This conference brought together researchers that were employing the then newly emerging digital computer technology in disciplines as different as psychology, computing, linguistics, neurobiology and engineering.

At the time, Behaviourism was at its peak in psychology. Behaviourism had arisen out of a partially justified methodological skepticism towards introspectionism in psychology, whose data was not observable by anyone but the introspecting subject and thus did not meet the scientific standards of the natural sciences. Therefore, the behaviourists demanded to confine scientific inquiry to physically measurable behaviour. The most radical critics went as far as to claim that mind and mental phenomena "could not be shown to exist and were therefore not proper objects of scientific inquiry at all" (Stilling, Weisler, Chase, Feinstein, Garfield, & Rissland, 1998, p. 335) and the very use of mentalistic language was, as a consequence, frowned upon.

The analogy between computing processes in digital computers (or formal Turing Machines, TMs) and the human mind drawn by the researchers in the newly founded discipline AI, therefore, fell on fertile grounds with researchers that were interested in studying mental phenomena. Digital computers perform intelligent tasks that previously only humans could do, such as logical reasoning, mathematical computation, syntactically correct chaining of words, etc. If we can physically explain and formally and functionally describe how the machine does it, why would the same not be possible for the human mind, the 'black box' of Behaviourism? Computer technology and AI provided the language and concepts that, in the oppressive scientific climate at the time, made it acceptable to use mentalistic terms without falling subject to accusations of lacking scientific rigour.

The Science of Cognition, rather than the Science of 'just' behaviour, therefore, is frequently defined in terms of this metaphor of the digital computer for the human mind (which comes in different variants, e.g., physical symbol system hypothesis (Newell & Simon, 1963); computational theory of mind (Fodor, 2000); cognition as computation or information processing (Stilling et al., 1998, p. 1)). It became the underlying dogma for the interdisciplinary study of the mind, in which cognitive psychologists and linguists empirically measure the behavioural data to be modelled; computer scientists and AI researchers generate the computational models of this data that map inputs to outputs and predict further not yet measured input-output mappings; brain scientists identify the neural circuits and brain areas that instantiate these formal models; philosophers take care of the mental side of things and relate the formal and scientific results to mind, which is scientifically not measurable. Had it worked, it would have been a great idea.

The problem with the mind-as-machine metaphor is that neither the human mind nor the human brain are very much like digital computers. A digital computer is a device that maps input

symbols to output symbols following syntactic rules, and, even though humans are much better at performing such mappings than most animals, it is by far not everything they do. Computers can model those aspects of our behaviour that are syntactic *in their nature*, but such behaviours are but a very limited subset of the things we do. Consequently, over the last 50 years of AI research, computer modellers have repeatedly run up to the limits of this metaphor. This led to the identification of a whole catalogue of problems that can ultimately be traced to originate from the mind-as-machine metaphor. A non-exhaustive listing features: 1.) the frame problem in AI: how to keep track of all relevant changes in the world? (Russell & Norvig, 1995) 2.) The credit assignment problem in machine learning: if I get positive feedback, what was it that I did right? 3.) the symbol grounding problem in philosophy: how do symbols get their meaning? (Harnad, 1990) 4.) the binding problem in neuroscience: selecting the features to build up the internal representation of the external world, how do I put back together correctly what I separated in the process of parsing? 5.) The problem of indexicality in formal semantics: how do I functionally derive word meanings that depend on context? The list could go on.

All these problems are a consequence of having separated the symbolic representational token from its meaning, a separation which characterises computational systems. Local structures do their job, applying syntactic rules without knowing if they are playing chess or launching a nuclear bomb. This ignorance of the algorithm is beautifully illustrated in Searle's famous 'Chinese room' thought experiment (1980), which features a Chinese interpreter that applies the rules of Chinese language without knowing any Chinese.

This is just one, and perhaps the most drastic implicit premise contained in the mind-as-machine metaphor. A number of assumption about brain and mind that are not supported by empirical evidence piggyback on this premise - assumptions that have been vehemently criticised over and over in the 50 year old history of AI. Other related assumptions, such as the idea that exact timing does not matter, that the brain/mind is functionally modularised, that cognition is the passive waiting for inputs, that there is an external world of objects, waiting to be represented and that unexplained homunculi provide meaning wherever it is lacking, are discussed later on in this chapter, throughout this thesis and in many of the references given.

However, the point of this section is not in the first place to convince the reader that there are problems with the mind-as-machine metaphor. The 'failure of AI', as it is commonly called, is, by now, acknowledged even by some of the most central and fundamentalist figures in Cognitive Science (e.g., Fodor, 2000). However, the methods, and with them the language, the concepts, the modelling assumptions and the rejection of other ways of doing Cognitive Science persist. Having started as a rebellion against the constraints that Behaviourism imposed on language, thought and action, computationalist Cognitive Science has now itself become an intellectual straightjacket, an obstacle in the way of scientific progress and the understanding of mind. While the mind-as-machine metaphor provided the language to describe cognitive processes that are syntactic in their very nature, it did not provide the language to talk about semantics, about meaning. This is a problem, because mind is an inherently meaningful phenomenon. Even worse maybe, the metaphor took away the language to talk about behaviour or anything external to the former black-box of Behaviourism, because it presumes that internal representation and symbol manipulation, the formal description of the mind-machine, is all there really is to know.

My personal belief is that both Behaviourism and computationalist GOFAI Cognitive Science have been so successful because they are based on seductively simple ideas. Is it time to replace computationalism with another seductively simple idea? I do not think so. The world is complex, mind is complex, the brain is complex, the body is complex. Any simple theory will be doomed to follow the same destiny, i.e., to rise, to turn into dogma, and to fall, but not to explain cognition. Fortunately, the enactive approach is not simple. Section 2.3 tries to capture the essence of what this new and still dynamic and evolving approach takes from different predecessors, some ancient, some more recent, and how it aspires to explain mind. But before that, I want to review a number of other proposed alternatives, some of which I consider more reasonable than others.

2.2 Alternative Paradigms

Sceptics have pointed out the above summarised limitations over and over again. But does giving up on computationalism imply giving up on the idea to scientifically explain mind and cognition? Or are there ways to cut out the mind-as-machine metaphor but to keep Cognitive Science as such an interdisciplinary project? Many proposals have been made to substitute the mind-as-machine metaphor with a new and different paradigm to be programmatic for a new Cognitive Science.

There is a tendency to conceive of such alternative proposals as a unified 'opposition', rather than as the diverse set of paradigms that it is. For instance, in *Connectionism, artificial life, and dynamical systems: New approaches to old questions*, Elman (1998) presents three alternative paradigms to the computationalism and describes how he believes they go hand in hand:

> "The three approaches share much in common. They all reflect an increased interest in the ways in which paying closer attention to natural systems (nervous systems; evolution; physics) might elucidate cognition. None of the approaches by itself is probably complete; but taken together, they complement one another in a way which we can only hope presages exciting discoveries yet to come" (Elman, 1998).

Earlier on, Elman writes

> "While there are significant differences among these three approaches and some complementarity, they also share a great deal in common and there are many researchers who work simultaneously in all three" (Elman, 1998).

In my opinion, Elman's take on the situation is misconceived. However, I do think it is quite a common misconception which is rooted in two facts that Elman observes: 1.) Alternative approaches tend to be driven by a common demand for more biological plausibility and 2.) there is methodological overlap between alternative paradigms. However, using the same methods and coming from the same route does not imply ideological agreement. Identifying the largest common denominator between different paradigms bears the danger of watering down the original radical and new proposals and dilute them "into a background essentially indistinguishable from that which they initially intend to reject" (Di Paolo, Rohde, & De Jaegher, 2008a). Therefore, I want to clarify the ideological commitments associated with some alternative paradigms that are all subsumed under the umbrella term *new AI*.

2.2.1 Connectionism

Connectionism is frequently conceived of as the most important alternative proposal to the classical paradigm. This perception is probably to do with the fact that Connectionism had been posed as an explicit challenge quite early on (McClelland, Rumelhart, & Hinton, 1986) and that Artificial Neural Network (ANN) theory had been developed alongside logics-based AI. Connectionism (or parallel distributed processing, PDP) proposes to replace GOFAI's digital computer with "a large number of simple processing elements called units, each sending excitatory and inhibitory signals to the other units" (McClelland et al., 1986, p. 55). Benefits of this approach are its "physiological flavour" (McClelland et al., 1986, p. 55) because ANNs are inspired by neuroscience, drawing the analogy between processing units and biological neurons. A lot of the paradigmatic debate in Cognitive Science has focused on identifying the differences between Turing Machine/logics based approaches and ANNs and their implications (noticeably: Fodor and Pylyshyn's (1988) conceptual criticism and the responses it triggered; Minsky and Papert's (1969) formal proof of limited computational capacities of perceptrons).

From an enactive perspective, ANNs are only of limited conceptual interest. In their non-dynamic form (i.e., feed-forward networks), they only represent input-output-mappings just like computationalist models. In their dynamic form (i.e., recurrent networks), they can represent dynamical systems - however, that a dynamical system takes the form of an ANN rather than just any differential equation is not of explanatory importance either. As argued extensively elsewhere (e.g., Cliff, 1991; Harvey, 1996), Connectionism suffers from most of the problems associated with the computationalist paradigm. Indeed, it is just a variant of the computational paradigm, not presuming 'cognition as digital information processing' but rather 'cognition as parallel distributed processing'.

ANN theory has produced some very useful formal tools, learning algorithms and representations for dynamical systems and mathematical functions. At its interface to theoretical neuroscience it has also generated models that contribute to the understanding of brain physiology and dynamics. This is as far as its significance goes. In order to understand mind, cognition and behaviour, it is necessary to investigate not just what comes in and what comes out, but much rather what happens in closed loop interaction with the world and how such physical agent-environment interactions relate to experience. ANN theory is not at the heart of such a project, it is not even an essential component.

2.2.2 Dynamicism

The dynamical hypothesis in Cognitive Science (van Gelder, 1998; Port & van Gelder, 1995) is a more recent alternative paradigm proposed which claims "that cognitive agents are dynamical systems" (van Gelder, 1998, p. 615). The problem with this approach is, again, that a mathematical formalism to substitute GOFAI's Turing Machine is proposed, rather than to part with the idea that a formal tool has to be at the core of Cognitive Science in the first place.

Dynamical Systems Theory (DST) does play an important role in the enactive approach, and I elaborate on this methodological importance in section 3.2 of the following methodological chapter 3. At this point, I only want to mention an example of the kind of model that is not enactive but falls within the realm of Dynamicism, in order to illustrate where the methods of Dynamicism

and Enactivism depart, despite the enormous overlap, which is much larger than the overlap with Connectionism.

Elman (1998) gives as an example of a DST model a recurrent neural network that is trained to recognise the context-sensitive formal language $a^n b^n$, which Elman sees as an example of a dynamical model of "realms of higher cognition" because it is "applied to the case of language" (Elman, 1998, p. 30). In the light of the previously identified problems with the computational paradigm, it is mysterious to me what this completely disembodied model which basically represents a pushdown automaton can explain about cognition: why is this model superior to a TM recognising the same formal language, or how does it not fall victim to the same criticisms? This is just to illustrate that the answer cannot be in the appropriate choice of formalism alone.

But there is a second part to why, even though I would never deny the importance of DST for Cognitive Science, I think Dynamicism is limited as a paradigm: a formal model in Cognitive Science cannot explain but an aspect of the *explanans*, it cannot itself be the phenomenon. I outline my ideas about the role of formal modelling in Cognitive Science in the following chapter (section 3.3).

2.2.3 Cybernetics, ALife, Behaviour Based Robotics (BBR)

While Connectionism and Dynamicism focus their criticism of the computationalist approach on the properties of the formalism used for modelling, both Behaviour Based Robotics (BBR, e.g., Brooks, 1995) and Artificial Life (ALife, e.g., Langton, 1997) emphasise the importance of embodiment and situatedness of cognition. The computationalist paradigm focuses on what comes in and what goes out but fails to account for how what goes out impacts in turn on what comes in (i.e., the *closure of the sensorimotor loop*) and its relevance for explaining cognition.

Associated with these approaches is a strong skepticism of the objectivist assumption implicit in computationalism, i.e., that the brain builds an internal representation of the external world which justifies to exclude the world itself from the explanation of cognition in favour of a Cartesian theatre (or as Brooks puts it: "the world is its own best model" (Brooks, 1995)). This skeptical position is frequently called *anti-representationlist*, even though I am not aware of anyone adopting this label for themselves. Harvey (1996, also personal communication at several oral presentations), however, appropriately remarks that from being the 'billiard balls' of explanation in computationalism (i.e., part of the *explanans*), human capacity to represent becomes an *explanandum* in non-computationalist paradigms and, though intriguing, loses its central role in explanation. He also points out that there are very different and partially contradictory meanings associated with the term 'representation' in Cognitive Science and everyday life (e.g., correlation, stand-in, re-presentation, something mental, something in the brain, a computational token, ...) and that computationalists are frequently reluctant to define their usage of the term which implies that it is a problematic term to use because of its fuzziness and the potential for misinterpretation it bears. I subscribe to this view.

Both BBR and ALife emphasise the fact that living organisms differ in that respect from digital computers; ALife can be seen as a direct counter-proposal to AI as in GOFAI that focuses on explaining "life as it is and how it could be" (Langton, 1997) rather than 'intelligence' which is associated with logics, rationality and the kinds of things that computers are good at. These

synthetic approaches clearly have their predecessors in the cybernetics movement during the first half of the last century (e.g., Ashby, 1954; Ryle, 1949; Holland, 2002, Holland on Walter's work from the 1940s/1950s), whose aim can maybe be identified as explaining living organisms as machines (not as Turing Machines (!)). Brooks sees his BBR approach in direct succession to the cybernetics movement, whose limitations he diagnoses to be due to the limited technologies and formal tools available at the time (Brooks, 1995). There is a multitude of opinions within the BBR and ALife community about several issues, such as whether simulation models count as embodied and situated. As a common denominator, BBR and ALife, in continuation of early cybernetics ideas, is to propose that behaviour has to be modelled/built in closed loop agent-environment interaction.

In my opinion, there is no contradiction between this paradigm and the enactive approach, as elaborated in section 2.2.5.

2.2.4 Minimal Representationalism/Extended Mind

There have been a number of proposals that explicitly aim at reconciling the old computationalist paradigm with the growing group of critics becoming aware of the need to take embodiment, situatedness and real-time interaction dynamics seriously. As we assess in (Di Paolo et al., 2008a),

> "In the opinion of many, the usefulness of enactive ideas is confined to the 'lower levels' of human cognition. This is the 'reform-not-revolution' interpretation. For instance, embodied and situated engagement with the environment may well be sufficient to describe insect navigation, but it will not tell us how we can plan a trip from Brighton to La Rochelle. [...] For some researchers, then, the usefulness of enactive ideas is confined to the 'lower levels' of human cognition. As soon as anything more complex is needed, we must somehow recover newly clothed versions of representationalism and computationalism"

Main proponents of this kind of approach include Clark and Grush (Clark, 1997; Clark & Grush, 1999) and Wheeler (2005). As we argue in (Di Paolo et al., 2008a), these proposals aim at incorporating syntactic symbol manipulation processes into an embodied and situated story in order to account for high level human reasoning. This approach is thus to abstain from the chauvinism associated with traditional computationalism (i.e., that a TM description will give you the whole story). However, such proposals extend the computationalist program, rather than to fundamentally change it: there will be some need to refer to dynamical, bodily and environmental variables, but at some level, cognition is and has to be still a homuncular symbol manipulation process working on internal representations. Those cognitive capacities presumed to be thus implemented are called "representation hungry problems" (Clark, 1997).

The model of value system architectures presented in chapter 5 illustrates some of the conceptual problems associated with hybrid architectures and homuncular modules. Problematic though these proposals may sound, they have to be taken seriously because they point towards the main challenges for an enactive Cognitive Science. There are, at present, not many enactive accounts of cognitive activities that involve the use of symbols, like language, mathematics or planning. Dynamical systems accounts frequently focus on cognitive capacities that are strongly rooted in the here-and-now, which leads cognitivists to believe that this is all these accounts can offer. Anthropological work on language (e.g., Núñez & Sweetser, 2006; Lakoff & Johnson, 2003) or

mathematics (e.g., Lakoff & Núñez, 2000) takes first steps to fill this gap. However, from the domain of computational modelling, there have been little contributions towards explaining the mentioned cognitive phenomena.

As outlined in section 2.4 below, for the enactive approach, this gap is not a failure but a challenge. There are no logical limits towards explaining these kinds of phenomena, like there are for the computationalist paradigm, but much rather horizons towards which this young paradigm can venture out next. For the present purpose, it is only important to point out that such hybrid or 'on the fence' positions are not variants of the enactive paradigm but, if at all, variants of the computationalist paradigm, in suggesting that human symbolic reasoning has to be a minimal form of symbolic digital computation.

2.2.5 Methodological Overlap, Ideology Worlds Apart

From the previous summary, it is easy to understand how alternative paradigms can get shuffled up: the shortcomings they aim to mitigate and the methods they propose overlap remarkably. However, as concerns the science-theoretic side of things, there are important differences and even contradictions between all these paradigms.[1] Most of them put too much emphasis on the descriptive formalism, just as computationalism does.

My position within the described landscape is to acknowledge a significant methodological overlap, but reject most of the labels just mentioned. For instance, even though I use both ANNs and DST as formal tools, I would not label myself a Dynamicist, and even less a Connectionist, because descriptive formalisms are not central to the enactive paradigm, which goes far beyond formal issues, whereas they are at the core of both Connectionism and Dynamicism. As concerns ALife as a paradigm *for AI*, I am more happy to label myself that way: I think ALife's closed-loop modelling approach is the way forward for modelling the kinds of phenomena I am interested in. I would, however, like to add the disclaimer that it is not my paradigm of choice *for Cognitive Science*. Even though the enactive paradigm in Cognitive Science has a space for synthetic methods in which my ALife simulation modelling fits, I do not think modelling or synthetic recreation are central to Enactivism. I elaborate on the status of formal tools and methods within my research in the following methodological chapter (section 3.3). The following section gives an account of the enactive approach as an alternative paradigm for the scientific study of cognition.

2.3 The Enactive Approach

The term 'enaction' in the context of cognition is usually associated with the publication of 'The embodied mind' (Varela, Thompson, & Rosch, 1991) and Francisco Varela et al. as key proponents, even though the term has been used in related contexts before (cf. Di Paolo et al., 2008a, section 2). I see myself very much in the tradition of the interdisciplinary research program put forward by Varela, which may be construed as a kind of non-reductive naturalism, emphasising the role of embodied experience, the autonomy of the cogniser and its relation of co-determination with its world. In this section, I outline my interpretation of the enactive approach, which mainly

[1] At least as they are phrased by their most radical proponents; I believe that many researchers applying the mentioned methods and labelling themselves accordingly may be a little more modest or less chauvinistic about their choice of method.

recapitulates the positions put forward in (Di Paolo et al., 2008a).

As dissatisfaction with the classical computationalist paradigm grows, the term 'enactive' gains in popularity. In the light of the paradigmatic confusion sketched in the previous section 2.2, there is a clear danger that the enactive approach as a paradigm is watered down, becomes a meaningless umbrella term or falls victim to self-contradiction. Therefore, the ideological commitments characterising this approach have to be made explicit. However, as the enactive approach is still emerging and developing, it is also important to avoid simplification, reduction and rushed exclusion of promising routes towards an open future. In (Di Paolo et al., 2008a), we write

> "In trying to answer the question 'What is Enactivism?' it is important not to straightjacket concepts that may still be partly in development. Some gaps may not yet be satisfactorily closed; some contradictions may or may not be only apparent. We should resist the temptation to decree solutions to these problems simply because we are dealing with definitional matters. The usefulness of a research programme also lies with its capability to grow and improve itself. It can only do so if problems and contradictions are brought to the centre and we let them do their work. For this, it is important to be engendering rather than conclusive, to indicate horizons rather than boundaries" (Di Paolo et al., 2008a).

The collection (Stewart, Gapenne, & Di Paolo, 2008) in which the cited contribution appears is an important step towards such an 'emancipation without dogmatisation'.

We have identified five central and conceptually intertwined concepts that constitute the core of the theory of enaction (Varela et al., 1991; Thompson, 2005), i.e., autonomy, sense-making, emergence, embodiment and experience, five ideas that partially imply each other.

2.3.1 Autonomy

The term 'autonomy' means that you live by your own ('auto') laws ('nomos'). The theory of autopoiesis (Maturana & Varela, 1980) argues that living organisms are autonomous because they constitute and keep building themselves and maintain their identity in a variable environment. This means that, at some level of description, the conditions that sustain any given process in a network of processes are provided by the operation of the other processes in the network, and that the result of their global activity is an identifiable unity, as it is best exemplified by the autonomy of the living cell.

Three things are important to realise about this idea of biological autonomy. 1.) The recognition of the agent as constructing, organizing, maintaining, and regulating sensorimotor interaction with the world is in direct opposition to a representationalist perspective in which agents mechanically represent and react to objects in a world with a pre-given ontology of meaningful objects. 2.) The constraints imposed on self-maintaining processes of identity generation are of *mechanical* nature; living organisms are bound by the laws of physics but the possibilities to re-organise themselves and, with them, the world of meaningful interactions they bring forth, are open-ended. This open-endedness contrasts with explicit design in computationalist approaches. Even if machine learning is a vivid field of investigation, such algorithms are *functionally* constrained by in-built laws. 3.) Against a common prejudice, autonomy does not equate to maximal moment to moment independence from environmental constraints (e.g., Bertschinger, Olbrich, Ay, & Jost, 2008; Seth,

2007). It means, contrariwise, "being able to set up new ways of constraining one's own actions" (Di Paolo et al., 2008a).

The living cell may be the best example for biological autonomy, but it probably is not the best example for the importance of autonomy in the scientific study of cognition. The cognitive capacities of cells, if you want to call them cognitive at all, are very limited. I elaborate on how autonomous identity preservation can happen at many possible levels, not only on the metabolic level, in section 5.1.1 (cf. Varela, 1991, 1997). Against another common prejudice, the enactive approach is not obsessed with bacteria cognition; Varela's late work was much more centred on the investigation of neuro-cognitive autonomy and cognition (e.g., Varela, 1999; Rodriguez, George, Lachaux, Martinerie, Renault, & Varela, 1999), and there are recent and interesting proposals that self-sustaining metabolism is altogether insufficient to give rise to mind or intentionality, which instead is postulated to result from self-sustaining closure at the behavioural level ('Mental Life'; cf. Barandiaran, 2008). Such contemplations of neuro-cognitive identity and autonomy are immaterial of the question whether or not such 'Mental Life' could exist without an organismic metabolic substrate, which is an open research question.

2.3.2 Sense-Making

The concept of sense-making is closely related to the concept of autonomy - it emphasises the constructivist and epistemological component in the enactive approach. In so far, "Enactivism differs from other non-representational and dynamical views such as Gibsonian ecological psychology (Varela et al., 1991, p. 203f). For the enactivist, sense is not an invariance present in the environment that must be retrieved by direct (or indirect) means. Invariances are instead the outcome of the dialogue between the active principle of organisms in action and the structure of the environment" (Di Paolo et al., 2008a).

As John Stewart remarked (in a plenary discussion at ARCo2006 in Bordeaux): the problem with information is not that there is not enough out there; the problem is that there is too much of it. There are infinite, countless invariances that could be detected and represented. Those *relevant* to the cogniser are those that are perceived, and what is relevant depends on the cogniser's organisation. The formation and perception of concepts, in turn, can alter the autonomous organisation of the cogniser, which can lead to the construction of new or destruction of existing meanings. Cognition therefore is a *formative* activity, not the extraction of meaning as if this was already present.

To realise this constructive role of the cogniser helps to disarm another common accusation, which is that constructivist approaches are non-naturalist or deny the existence of the physical world. The only thing denied is the observer-independent existence of meanings and secondary qualities - not the existence of a universe of meaningless matter and physical constraints as such.

2.3.3 Emergence

In order to illuminate the concept of emergence, I want to return to the example of the living cell. How do we know the cell is alive? And what exactly is alive? "The property of continuous self-production, renewal and regeneration of a physically bounded network of molecular transformations (autopoiesis) is not to be found at any level below that of the living cell itself." (Di Paolo

et al., 2008a). It seems ill-conceived to call any of the component parts, a protein, the DNA strands, etc. alive: these are just physical structures, the material substrate of the living cell that is constantly changed and renewed. It is undeniable, however, that the phenomenon of life is as real as it could be.[2]

We can very well scientifically investigate the material substrate of the living, and how it brings about relational properties such as 'life', 'death' or 'survival', without ever being able to (or wishing to) reduce them to the physical substrate. In the same sense, we can scientifically study the physical processes from which mind and meaning emerge. The latter are then not to be reduced to physical components of either the agent or its environment, but belong to the relational domain established between the two.

The centrality of the concept of emergence is at the root of enactive skepticism towards functional localisation as it is practiced in traditional cognitivist psychology, AI and neuroscience. The problem is not that there would be insufficient evidence or not for such assignments, but much rather that this kind of reductionist assignment is a category mistake. This question is explored further in chapter 5.

2.3.4 Embodiment

Embodiment is a concept widely discussed and valued in Cognitive Science. Therefore, I do not want to dwell too much on the dated idea that cognition was the meat in the 'classical sandwich' (Hurley, 1998) (i.e., squashed between the negligible peripheral sensor and motor systems that generate symbolic representations and execute symbolic outputs as actions).

Instead I want to point out the important difference between embodiment and mere physical existence. "[A] cognitive system is embodied to the extent to which its activity depends non-trivially on the body. However, the widespread use of the term has led in some cases to the loss of the original contrast with computationalism and even to the serious consideration of trivial senses of embodiment as mere physical presence - in this view a word-processor running on a computer would be embodied, (cf. Chrisley, 2003)" (Di Paolo et al., 2008a). Embodiment is not 'symbol grounding' (Harnad, 1990) through implementation, an idea that keeps up the Cartesian separation between cognition and 'reality'. Much rather, embodiment means that cognition *is* embodied action, in that the sensorimotor invariants our body affords in interaction with this world constrain and shape the space of meanings constructed.

2.3.5 Experience

Steve Torrance (personal communication) remarked that experience is an 'embarrassment' for the computationalist approach: a full blown cognitivist architecture, which supposedly explains cognition, fails to account for one of the most central feats of the mental, i.e., what it feels like. With decades having passed since Behaviourism, the 'c-word' (consciousness) has become less and less of a taboo even in mainstream Cognitive Science. What it feels like has become one of the most important topics of debate and controversy in the philosophy of mind ('explanatory gap' (Levine, 1983) and qualia debate as the Cognitivist variant of the mind-body-problem). It is important

[2]Even if this seems to be forgotten by some modern biologists, as Stewart (2004) argues, which is quite ironic, given that biology is the study ('logos') of the living ('bios').

to realise that the way this debate is led from within the computationalist paradigm is Cartesian (or closet Cartesian), in that the mental is considered a different kind of thing from anything else (objects, the world, meaning, the brain, representation, symbol manipulation; anything 'real' and physically explainable) and we are therefore left with the impossible and artificial task to re-unite these two things that we tore apart.

The enactive approach does not deny that experience does not manifest itself as physical objects. But in not being matter, experience is in good company, with other non-material, non-reducible, but nevertheless real phenomena such as life, meaning, emotions or intentionality. "Experience in the enactive approach is intertwined with being alive and enacting a world of significance" (Di Paolo et al., 2008a), not just as data to be explained, but as a guiding force in research methodology. This is not to say that the study of experience (through scientific or non-scientific means) is not methodologically problematic. Personally speaking, I find experience the most difficult factor to incorporate into a paradigm for the Cognitive Sciences. But it surely does not help to pretend experience does not exist.[3] Section 3.5 discusses in more detail how experience can be methodologically incorporated in Cognitive Science, discussing the distinction between first, second and third person approaches.

A last issue to be clarified is the apparent contradiction between the centrality of the concept of experience in the enactive approach, on the one hand, and, on the other hand, its strong interest in non-human life and cognition and the phylogeny of cognition. Through experience, we know what things mean to us, to our socio-linguistic selves. How can we say anything meaningful about the meaning space of a different species, with a different or more primitive organisation, who cannot even linguistically express themselves? We can find the answer in Jonas' (1966) work and Weber's (2003) extension of it: the 'ecstatic' character of the living allows us to understand, from organism to organism, what something means to another subject, not as 'what it feels like', from the inside, but as 'what it means', reading the signs.

> "the patient who is not anymore able to articulate himself, animals, even a paramecium that cramps before it is killed by the picric acid dribbled under the cover slip, the sadening look of a limb plant, the phetus that defends itself with hands and feet against the doctor's instruments - they all *present* the meaning of what is happening to them" (Weber, 2003, p. 118).[4]

2.3.6 The Roots

This brief outline of the enactive approach and its central concepts and ideas has made little reference to the numerous predecessors from many scientific disciplines or related contemporary currents of research. It is important to acknowledge these sources of inspiration and explain where the enactive approach came from.

Maybe the most important predecessor is Maturana and Varela's own theory of autopoiesis (e.g., Maturana & Varela, 1980, 1987). The idea of autopoiesis as the organisation of the living

[3] Note (23.02.2008): To some people, this is not as bizarre a suggestion as it seems. Only the day before yesterday, I have again been told exactly that (i.e., that mind does not exist) by a member of Sussex faculty, who suffered a sudden outburst of scientism.

[4] My translation: "der nicht mehr artikulationsfähige Kranke, Tiere, ja sogar das Pantoffeltierschen, das sich zusammenkrampft, bevor es von der unter das Deckglas geträufelten Pikrinsäure getötet wird, der trauig stimmende Anblick einer welken Pflanze, der Fötus, der sich gegen die Instrumente des Arztes mit Händen und Füßen wehrt - alle *zeigen* die Bedeutung dessen, was ihnen widerfährt" (Weber, 2003, p. 118).

still plays an important role in the enactive approach (previous sections) but, as I conceive it, autopoietic theory is more concerned with theoretical and epistemological questions, whereas the enactive approach focuses on scientific practice and explanation. Also, with the idea of 'enaction as embodied action', the enactive approach emphasises the active and engaging side of knowledge construction, whereas the original formulation of autopoietic theory has sometimes (unjustly) been criticised to endorse solipsism or non-naturalism.

There are, of course, also numerous predecessors and contemporary researchers with large ideological and methodological overlap among the countless participants in the universal and millennia old pursuit to explain mind. In section 2 of (Di Paolo et al., 2008a), we provide a non-exhaustive listing of scientific currents that relate to the enactive approach, featuring, e.g., Piaget's theory of cognitive development through sensorimotor equilibration (Piaget, 1936), the philosophical strands of existential phenomenology, continental biophilosophy and American pragmatism, holistic dynamical systems approaches in neuroscience, ALife researchers in AI and Robotics, etc. It is important to realise the cognation between these predecessors and related approaches and the enactive approach, not just to get a better impression of what enaction is all about, but also because the insights and findings resulting from such approaches can be used to enrich and advance an enactive understanding of the mind. Throughout this dissertation, I refer to such related work as a complement or source of inspiration.

2.4 Challenges, Criticisms and Simulation Models

Being a passionate promoter of the enactive paradigm, I want to use this section to identify how this young paradigm can and should be growing, both in general and as concerns my own work. As already pointed out in section 2.2.4, there are areas in enactive Cognitive Science that are still underdeveloped. In particular, those are the GOFAI strongholds in which proponents of minimal representationalist views postulate "representation hungry" (Clark, 1997) problems that require explicit symbol manipulation processes for their explanation. Most of these involve higher levels of cognitive performance: thinking, imagining, engaging in complex interactions with others, and so on. There is no reason to believe that the enactive approach is not able to explain these kinds of phenomena, but as long as it fails to do so, skeptics cannot be hushed.

In (Di Paolo et al., 2008a), we argued that "[w]e must not underestimate the value of a new framework in allowing us to *formulate the questions in a different vocabulary*, even if satisfactory answers are not yet forthcoming". In order to prove this point, we give examples from different areas and from our own modelling work (for instance the models presented in chapters 5 and 6 of the current dissertation). This dissertation can be seen to pursue a very similar objective, even if in much more depth and with a much more focused methodological research question.

As already stated in chapter 1, working with ALife methods to investigate and build intelligent and adaptive behaviour, I quickly became frustrated by the apparent incapacity of these approaches to address questions of human level cognition. This frustration resonates with the previous point about computationalist strongholds and underdeveloped areas in enactive Cognitive Science.

A consequence of the mind-as-machine metaphor and the modularisation of function in the computationalist paradigm is that it is presumed that modelling increasingly sophisticated cognitive processes equates to adding more functional modules and computational complexity to the AI

model of cognition. This is not the same in enactive Cognitive Science. The models presented in this dissertation address five problems to be located in different disciplines and levels of sophistication in Cognitive Science. Yet all of them strive for minimalism of the simulation and for modelling the essence of the behavioural dynamics. With my work, I show that a complex *explanandum* does not require a complex model to form part of the *explanans*.

In terms of the problems addressed, the procession of simulation models here presented reflects my personal procession towards identifying the kinds of questions of high level human cognition that I believe the enactive approach will be able to address next, particularly with the subset of methods employed and developed by myself. A crucial moment in this procession was learning about the experimental work in Sensory Substitution/Perceptual Supplementation to explain the sensorimotor basis of space cognition (Lenay, 2003) conducted at the GSP in Compiègne. Space cognition, and in particular the conscious experience it involves, are clearly very abstract and high level cognitive capacities and the experimental approach taken elucidates the origins of spatial concepts and percepts. I used Evolutionary Robotics simulation modelling to complement this approach and, as a last step in this dissertation, identified an own question of interest, i.e., the sensorimotor basis of simultaneity perception and its relation to adaptation to sensory delays, as a question I would like to address from an enactive point of view, combining experimental, experiential and modelling work. The results from this project are just a first step in the investigation of this intriguing question. Performing this step and knowing which step to take next, however, can be seen in itself as a major result of this dissertation.

The route I have taken is just one in an infinite space of future possibilities for enactive Cognitive Scientists. "A proper extension to the enactive approach into a solid and mainstream framework for understanding cognition in all its manifestations will be a job of many and lasting for many years. [...] The strength of any scientific proposal will eventually be in how it advances our understanding, be that in the form of predictability and control, or in the form of synthetic constructions, models, and technologies for coping and interacting with complex systems such as education policies, methods for diagnosis, novel therapies, etc" (Di Paolo et al., 2008a). There are other challenges for the enactive approach that will require different methods. I explore, name and evaluate the usefulness and scope of applicability of Evolutionary Robotics simulation modelling for such future challenges throughout this dissertation. In particular, I identify interdisciplinary minimalist work on the sensorimotor basis of perception as a promising route to explain problems of high level human cognition and thus invade computationalist strongholds.

Chapter 3

Methods and Methodology

Due to the high methodological emphasis of the current thesis and the fact that it develops and presents a new methodological framework, this chapter contains a lot of novel material. Rather than to just iterate proven methods, large parts of this chapter are dedicated to science-theoretic and methodological argument, to explaining the methods I propose and to identifying their scopes and limits (hence the title: 'Methods and Methodology', rather than just 'Methods').

The first section 3.1 ties in with issues already raised in chapter 2 about the implications of a constructivist-enactivist world view, which denies the existence of an observer-independent reality, has for scientific explanation. In a similarly general style, section 3.2 assesses the importance and position of the mathematical language of dynamical systems theory for the enactive approach as I perceive it. Section 3.3 introduces Evolutionary Robotics simulation models. It presents technical details of the ER models used for the modelling parts of the current dissertation and discusses their role in scientific explanation. The following two chapters 4 and 5 present results from Evolutionary Robotics models that rely exclusively on the methods and methodological considerations outlined in this first part of the chapter.

Section 3.4 introduces Perceptual Supplementation as a minimalist experimental approach to human perception and sensorimotor adaptation, which is mainly based on ideas and methods developed by the Perceptual Supplementation Group in Compiègne. I realised the experimental parts of the interdisciplinary project (chapters 9 and 11) in collaboration with the Perceptual Supplementation Group, and the simulations presented in chapters 6, 7 and 10 model results from research in Perceptual Supplementation. An absolutely essential dimension of study of human cognition and perception is subjective experience. It is, however, the methodologically most difficult dimension. In section 3.5, I discuss first, second and third person approaches to the study of experience. This is the least developed part of the interdisciplinary framework I devise and, as I self-critically point out there and in the conclusion (chapter 13). Finally, section 3.6 brings together the methods presented to outline how they can be applied in mutual benefit, in particular Evolutionary Robotics simulations, minimal experiments on human sensorimotor adaptation and the study of rudimentary perceptual experience which is applied in the study on adaptation to sensory delays (chapters 8-12). This is the methodologically most novel part of this dissertation and I see it as one of its most important results.

3.1 The Scientist as Observing Subject

In the enactive view, knowledge is not represented, knowledge is constructed: it is constructed by an agent through its sensorimotor interactions with its environment, co-constructed between and within living species through their meaningful interaction with each other and, in its most abstract and symbolic form, knowledge is co-constructed between human individuals in socio-linguistic interactions (see also chapter 2 section 2.3 and 5 section 5.1 for an elaboration of ideas on values and meaning construction). Many of the ideas presented in this section have been previously published in (Rohde & Stewart, 2008).

Science is a particular form of social knowledge construction, characterised by certain rules, dogmata, procedures, objectives and *the use of formal languages and techniques of measurement*, which, if applied correctly, endow the thus generated knowledge with certain properties that make it somewhat special. Most noticeably for modern human society, the thus constructed knowledge can be taken further, following the rules of logic and mathematics and allows us to build powerful tools, machines and medicines, or to predict events beyond the scope of our perceptual or intuitive grasp of regularities in our environment.[1] The pragmatic power of (some) scientific knowledge, should, however, not seduce us to subscribe to some form of scientism, assigning scientific knowledge *ontological* privileges which it does not deserve. The significance of scientific knowledge always derives from the context of its generation and from what it means for an individual or a group of individual (e.g., a society), just like any other form of knowledge.

A constructivist Cognitive Science thus finds itself in a situation where the snake bites its own tail: it applies the rules and methods of science to explain processes of meaning construction, an *explanandum* that subsumes the application of the rules and methods of science itself. In their early work on autopoiesis, cognition and the principles of life, Maturana and Varela (1987, 1980) have crucially identified and discussed this status of the scientist as observer and what it implies for scientific practice in biology and Cognitive Science. Maturana's statement that "everything said is said by an observer" (Maturana, 1978) has become programmatic for the epistemological strand of radical constructivism in the 80s and 90s, and their writings have crucially influenced many pioneers of enactive and proto-enactive approaches in the Cognitive Science and biology (e.g., contributors to Varela et al., 1991).

Maturana and Varela, as well as their explicit followers, have phrased the implications of an observer-science as it is relevant to the questions addressed in the current dissertation most clearly. I thus refer mainly to those sources. It should, however, be noted that there are scientists and science theorists in and outside Cognitive Science (e.g., Bitbol, 2001; Kurthen, 1994) that equally recognise the constructivist nature of scientific activity and the necessity to include the observing subject into a scientific story as part of both *explanans* and *explanandum* at a time. A constructivist and non-objectivist science makes references to the specific processes of scientific knowledge construction where necessary (many of the issues raised in this section are relevant for the study of experienced simultaneity and come up again in chapter 8 on time).

For the case of Cognitive Science this is particularly tricky, as Cognitive Science is the science of mind and cognition, with science being an activity that is largely about measurements, obser-

[1] As Mike Leies once rightly pointed out in conversation: "It seems weird to sit in an airplane 3000 ft above the ground and say you do not believe in science".

vations and quantification, and mind and cognition being phenomena that are basically neither observable nor measurable nor quantifiable – which is what lead Descartes to his dualistic world view, distinguishing mind, the *res cogitans*, from basically anything else in the world which can directly take measurable effect in the environment and thus manifest in space, i.e., the *res extensa*. Whilst our access to objects and events is mediated through our sensorimotor interaction with the environment, our access to mental phenomena is direct and subjective, cognition manifests as *experience*, a category not usually considered part of the scientific program. Cognitive science thus has the thankless task to explain (amongst other things) the *qualitative* dimension of cognition, including the experience of emotions, intentions, colours, numbers, memories, insights, competencies, communication, etc. without actually having the words to express the *explanandum* in the first place.

The way traditional Cognitive Science deals with this problem is, typically, to *define* unmeasurable mental phenomena in terms of physically measurable variables and to *reduce* them to physical and quantifiable processes. Prominent examples of this practice include:

- The reduction of mind states to physical brain states on the basis of correlated occurrence, a practice that is popular with some philosophers of mind working in the qualia debate and on the neural bases of consciousness. In its most consequent and extreme form, this reductionism results in eliminativism (e.g., Churchland & Churchland, 1998).
- The functional reduction of cognitive phenomena to physically measurable processes that convincingly appear to bring about that cognitive phenomenon in an entity that is not oneself (Turing-test approaches, after Turing, 1950), a technique that is more commonly adopted in the areas of artificial intelligence and cognitive modelling and underlies Dennett's (1989) ideas on the 'intentional stance'.

The problem with these reductionist approaches is, in a nutshell, that by picking an isolated physical phenomenon and explaining it, you explain the isolated physical phenomena you pick, but not cognition.

Instead of indulging in ideological quarrels, which I did already in chapter 2, I want to argue how to scientifically study cognition, in a *bona fide* enactive way, avoiding the reductionist practices just mentioned. Firstly, it is necessary to establish exactly how measured empirical findings relate to experiential phenomena, rather than just to identify a local correlation and take it out of its physiological, physical and semantic context, such as when reducing mental states to brain states. Section 3.5 discusses the methodological difficulties associated with the study of experience in detail. Secondly, in order to be able to say something meaningful about functional aspects of cognitive faculties, it is important to explain the mechanisms that generate it, rather than just to explain some mechanism that successfully imitates particular aspects of the cognitive faculty under investigation. We have developed and discussed this point, focusing on the example of the scientific study of autonomy in (Rohde & Stewart, 2008) and the reminder of this section reproduces our argument.

The scenario developed by Alan Turing in his 1950 classic paper 'Computing machinery and intelligence' (Turing, 1950), which he called the 'imitation game' expresses a deep pessimism towards the possibility to properly scientifically account for intelligence or cognition. Via a language interface, what is tested is the capacity to trick a human being into thinking that it was interacting

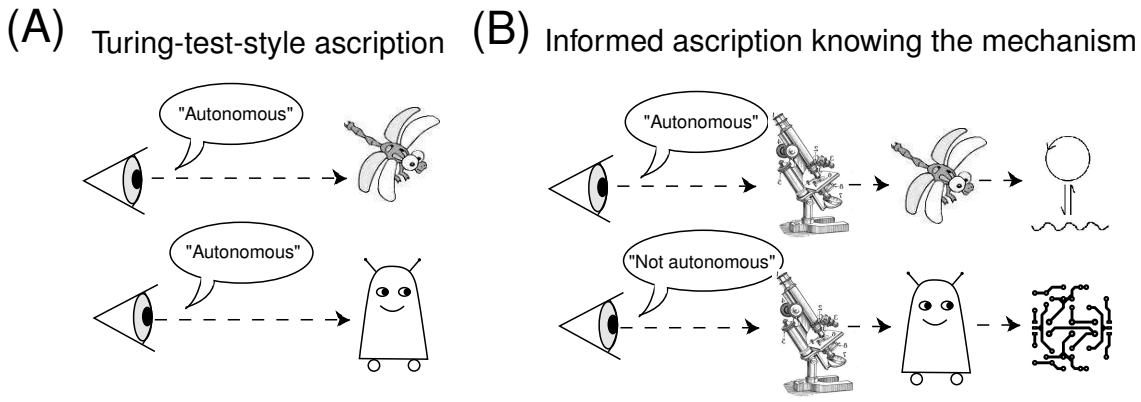

Figure 3.1: Illustration of ascriptional judgments of autonomy based on naïve observation (A) and scientific study of the generative mechanisms (B).

with another person, assuming that this capacity would presuppose some form of thinking in the machine. Turing's original formulation of the test was rather tame, i.e., that towards the end of the 20th century "an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning [a computer]" (Turing, 1950) and may even have approximately true: with simple techniques such as pattern matching, and psychological knowledge that allows to predict the most common questions asked in this kind of situation, some language programmers (not programs!) are quite good at tricking humans into the belief that they are actually communicating with a software agent, if only for a short while. The reality of how such systems are programmed and the kind of mistakes they make, however, reveals that these agents do not actually think or have any grasp of the meaning of the symbol strings they produce. This is what Searle (1980) illustrates in his famous 'Chinese room' thought experiment.

As we argue in (Rohde & Stewart, 2008), knowledge about the mechanisms that generate a phenomenon has a tendency to produce such reactions of disenchantment, the prime example being to know how a conjuring trick works, which clearly takes away the excitement about the seeming supernatural powers at work being profane sleigh of hand or visual illusions.

An important point to realise is that acquaintance with the underlying mechanism does not necessarily lead to disenchantment. On the contrary, sometimes, knowing how something works can produce the opposite effect: for example, a glider in the game of life does not look any different from a first-generation computer game sprite if you just look at it moving around on a two-dimensional grid. Only if you learn about the local cellular automata rules that underlie the emergence of a glider, their simplicity and the fact that they do in no way directly specify any of the emergent behaviour and appearance of the glider, it turns into a fascinating phenomenon, and there is no ulterior knowledge that I can acquire that could take this fascination away.

Applying these ideas to the study of cognition, our argument is that learning about the simple algorithms and rules of symbol manipulation that bring about seemingly intelligent or linguistic behaviour in GOFAI systems can leave behind a similar taste of charlatanry as the revelation of a conjurer's trick. I have personally experienced this disappointment many times with laymen when telling them how the robots work that they saw do impressive things (such as playing a violin or taking orders linguistically and execute them) in a short TV clip, from which naïve specators

presume that their capacities would generalise. Figure 3.1 illustrates this discrepancy in the case of ascription of autonomy to robots or living organisms: if autonomy (or any other cognitive capacity) is ascribed to a robotic agent using a kind of Turing-test that relies on superficial acquaintance in (A), knowledge about the generative mechanisms can lead to a revision of judgment in (B). When studying the autopoietic organisation of a living organism, acquaintance with the mechanism does not usually have this disenchanting effect.

It should be noted that I do not believe that being convincing is something inherent to living or ALife-style processes. Just as there are many genuinely fascinating machines (such as cars and computers), people can also get disappointed with processes generated by living organisms. For instances, there is a tendency to be disappointed by stigmergic processes in insects, as the example of the digger wasp (discussed, e.g., in Dennett, 1985) shows: the wasp appears to have an elaborate plan of clearing a tunnel it dug before putting a larvae in it. However, by dislocating its larvae while the wasp is inside the tunnel, the wasp can be trapped in an 'infinity loop' of repeatedly checking whether the tunnel is blocked. This reveals that it does not actually *know that* it is clearing the tunnel, in the sense of understanding the concept of tunnel clearing, but much rather *knows how* to clean the tunnel, following a sequence of behaviours that are triggered by changes in the environment. This behaviour is in crucial ways similar to a computer executing an algorithm and can lead to disenchantment in the same way.

We therefore propose in (Rohde & Stewart, 2008) to substitute a Turing-test style statistic measures of intuitive ascriptional reaction with informed ascriptions based on the scientific knowledge about generative mechanisms. This is not to propose a project of defining cognitive or mental faculties in terms of the physical properties of the processes that generate it or to engage in any other form of reductionist activity. It is proposing to make use of the beneficial characteristics that scientific knowledge has (as outlined above) in the larger endeavour to understand and explain mind and cognition, which is, in the end, what 'Cognitive Science' is all about. Apart from being more robust and reliable than many other forms of knowledge, scientific knowledge has the advantage that it is subject to inter-subjective debate and agreement, which can resolve controversies about whether or not a mechanism 'counts': "[if] the disagreement remains within the scope of a single paradigm, the normal process of Popperian refutation (or not) will lead to progress. If the disagreement occurs between incommensurable Kuhnian paradigms, then an element of subjective choice may remain" (Rohde & Stewart, 2008, see figure 3.2).

When we presented our ideas at the EU-cognition workshop on modelling autonomy in San Sebastián in March 2007, our argument was criticised by several researchers (most vehemently by N. Bertschinger) which compared knowledge about generative mechanisms with the hypothetical possibility of generating a perfect and complete grasp of the surface behaviour (i.e., how inputs and outputs relate over time). Supposedly, this 'LaPlacian Demon' type knowledge would be as powerful a basis for identifying autonomy as the scientific study of the generative mechanisms. Without even entering into a metaphysical quarrel whether or not this is strictly true in a principled way, this argument can be easily put to rest with epistemic arguments. Apart from the fact that for most real-life complex entities (and in particular living organisms), humans would be incapable of grasping the entirety of their sensorimotor couplings at once and confidently judge about their properties as a whole, the question to ask is: why bother, if we can as well study the generative

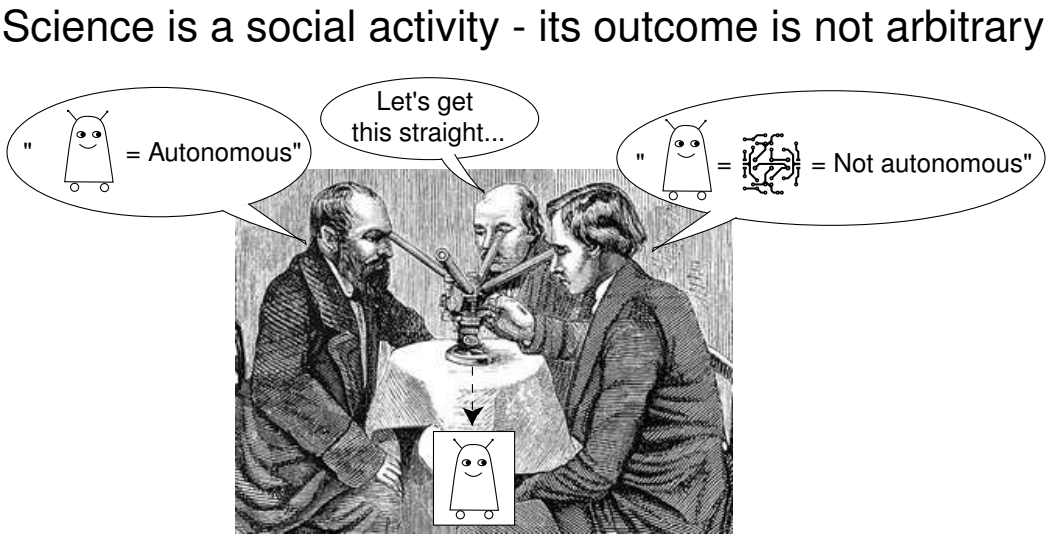

Figure 3.2: Illustration of the social dimension of scientific knowledge construction.

mechanisms?

3.2 Dynamical Systems Theory

3.2.1 Definition

In this section, I introduce some of the key terms and definitions in Dynamical Systems Theory (DST) that I refer to throughout this dissertation. The definitions here used stem from the following sources: (Strogatz, 1994; Rohde, 2003; Ross, 1984).

A state x of a dynamical system is a set of system quantities that allows the complete description of the system's development across time. Formally, a state is a variable assignment to a set of variables (state variables) of a dynamical system. In a dynamical system that models a real world system, the state variables correspond to measurable quantities. Apart from state variables, a system can have control parameters, which can change on a slower time-scale than the state variables. Their change is not accounted for in the description of the system: control parameters define a parameterised set of different dynamical systems.

Dynamical systems can either be given as a set of differential equations (time-continuous) or as a set of difference equations (iterated maps; time-discrete). In the current thesis, the dynamical systems investigated are differential equations, even if they are investigated discretised in computer simulation (see below).

I will not go into the details of different types of differential equations (ordinary, partial, stochastic, ...) and their formal properties here (see (Strogatz, 1994) for an accessible introduction). The only important concepts I want to briefly discuss at this point is the distinction between *linear and non-linear* dynamical systems and the notion of an *attractor*.

A linear dynamical system is basically a dynamical system in which the behaviour of the whole system is equal to the sum of the behaviours of its parts. This is in accordance with the general definition of a linear function in mathematics. In order for a differential equation to be linear, the terms that describe the change of the state variables must, therefore, not contain any non-linear

functions of state variables, such as power functions, products, trigonometric functions, etc. If they do, the differential equation is non-linear.

The mathematical tools for the analytical computation and analysis of non-linear differential equations are to the point not very advanced and require advanced mathematical skills. Therefore, computer simulations are important in the study of dynamical systems. In order to simulate time-continuous dynamical systems in digital computer simulation, the differential equations have to be discretised using numerical methods. In this thesis, the only numerical method used is the forward Euler method which approximates the change in state of a differential equation $\dot{x} = f(t)$ after a time step of length h as

$$x(t+h) = x(t) + hf(t) \tag{3.1}$$

Among the interesting properties of dynamical systems are what is called *attractors*. According to Strogatz, "there is still disagreement about what the exact definition [of an attractor] should be" (Strogatz, 1994, p. 324). I reproduce his definition here. An attractor is a closed set of states A that is *invariant, attracts an open set of neighbouring initial conditions* and is *minimal*.

'Invariant' means here that any trajectory that starts in A ends in A. Invariant sets can be fixed points ($A = x^*$ with $f(x^*) = 0$), limit cycles (circular orbits $\in A$) or strange (chaotic, fractal) sets, which "exhibit sensitive dependence on initial conditions" (Strogatz, 1994, p. 235), i.e., trajectories within A starting at states that are very close will describe very different orbits within A. Whilst fixed points can also exist in linear dynamical systems, limit cycles and strange attractors exclusively occur in non-linear dynamical systems.

The set of initial states attracted to A is called the *basin of attraction B* of an attractor and it is characterised by the fact that the distance from $x(t) \to 0$ as $t \to \infty$. An invariant set without a neighbouring basin of attraction is not an attractor. Such invariant sets are *unstable* or - in rare cases - *semi-stable*.

Minimalism means here simply that there is not subset of A for which the same properties (invariance, asymptotic stability) hold.

An orbit within the basin of attraction of an attractor that converges towards the invariant set is called a *transient*. A system is globally stable if all system states converge to a single attractor, it is multi-stable if it has more than one attractor. A convergent (dynamically trivial) dynamical system is one that has only fixed point attractors. A dynamical system is called an open system if it interacts with the environment; otherwise, it is called a closed system.

3.2.2 The Explanatory Role of DST

Being based on the 'Mind as Machine' metaphor, traditional Cognitive Science centres around a particular mathematical formalism, i.e., the Turing machine/automata theory as the fundament on which to built a unified interdisciplinary science. Some approaches that are critical towards classical computationalism and question the central role of this metaphor have tried to put other formal languages in its place, such as Connectionism proposing ANNs and Dynamicism proposing DST (cf. previous chapter, section 2.2). Van Gelder's (1998) proposal of the 'dynamical hypothesis in Cognitive Science' distinguishes the *nature hypothesis* and the *knowledge hypothesis* (van Gelder, 1998) as two sides of the same coin. The nature hypothesis is the hypothesis that what is cognitive about a cognitive systems is fully captured by an abstract formal description of its behavioural and

brain dynamics, i.e., it *is* this dynamical system, which can, in principle, be variably instantiated in material terms. The knowledge hypothesis is that a cognitive system is best studied with DST as formal tool.

The dynamical turn in Cognitive Science has gained in impact over the last years (e.g., Beer, 2000; Port & van Gelder, 1995; Thelen & Smith, 1994). Researchers identifying with Dynamicism work in areas as different as linguistics, physiology, cognitive psychology, developmental psychology, cognitive neuroscience, etc. Broadly speaking, the enactive approach can be seen as forming part of this dynamical turn, even though its core assumptions are not identical (cf. chapter 2). This difference does not entail a reservation: I am hugely sympathetic towards nearly all the work done under this label. I just want to stress that, in contrast to van Gelder's dynamical hypothesis, DST is not seen as a privileged formalism, but just a very suitable language for formalising the material aspects involved in cognition.

The reason why DST is the pre-dominant formal language of choice in enactive Cognitive Science are the same reasons that assign DST an important role in all natural sciences, and in particular in physics. As developed in chapter 2, the enactive approach investigates the mutual determination and constraining of the material or mechanistic level and the behavioural, cognitive and meaningful level. The enactive approach is concerned primarily with the origin, adaptive change and maintenance of invariant emergent structures. Such self-organisation is an inherently dynamical phenomenon. DST, as the language of physics, serves to describe the evolution of a whole situation over time, including an agent, its body, its environment and its brain. In order to describe and model embodied and embedded agents in a way that minimises prior assumptions about how structure relates to function, DST as a descriptive formalism has a clear competitive edge because of this capacity to describe physical processes in general. For the description and study of the mechanistic or physical level without building in prejudices about functionality of structure, DST suggests itself.

3.3 Simulation Models, Evolutionary Robotics and CTRNN Controllers

3.3.1 Technical Details of Evolutionary Robotics Simulations Presented

Evolutionary Robotics (ER) is a "technique for the automatic creation of autonomous robots [...] inspired by the [D]arwinian principle of selective reproduction of the fittest" (Nolfi & Floreano, 2000, preface). In this approach, some aspects of the robot's or simulated agent's architecture are specified, but others are underspecified. These are left to be determined in an automated way by an evolutionary search algorithm, according to the optimisation of an abstract performance measure called the 'fitness function' (see figure 3.3 for an illustration of the process).

In all the ER models presented in the scope of this dissertation, the parameters evolved are the parameters of the neural network controller and all experiments have been conducted in simulation. In this section, I present the algorithm and techniques that are common to the different models I present (control network and parameter ranges, genetic algorithm etc.). In each of the modelling chapters 4, 5, 6, 7 and 10, more technical details are provided that are specific to the model. In some of the models, there are deviations from the general principles described here. These deviations are pointed out within the modelling section of the respective chapter.

Apart from the model presented in chapter 4 all the simulation models rely fully on my own

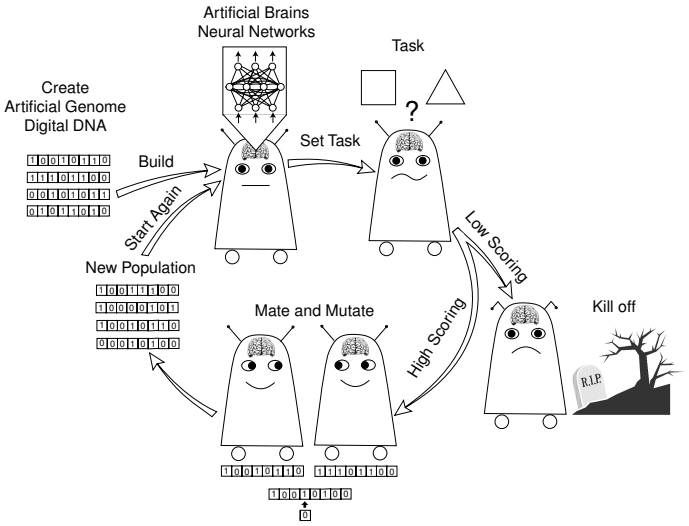

Figure 3.3: Illustration of the evolutionary cycle in Evolutionary Robotics.

code and have been implemented in Java. The general class structure adopted is to have distinct classes for the GA (in which the artificial genotypes are stored and managed), the individual (in which the genotype is interpreted in terms of neural parameters and the network dynamics are computed) and the simulation (in which the motor output from the networks are interpreted and the environmental dynamics, the fitness and the sensory states are computed). The GA calls the individual and the individual calls the simulation. Data recorded from simulations is

- The fitness values of all agents in all generations (to analyse evolvability).
- The genotype of the best individual in each generation (to analyse transitions in behaviour over evolution).
- The genotype of all individuals from the last generation (to seed a continued evolutionary run, if required, or to check for behavioural diversity)

Additionally, there were classes to test evolved behaviour under variable circumstances and to record and visualise sensorimotor trajectories, environmental variables and the state of the controller.

Continuous-Time Recurrent Neural Networks (CTRNNs)

A method used and promoted by Beer is the use of a particular network type for ER neural control, i.e., Continuous Time Recurrent Neural Networks (CTRNNs, e.g., Beer, 1995). Even though the dynamical properties of CTRNNs can be seen as idealisations of real neural dynamics, I do not use CTRNNs as direct analogies for the brain or brain areas. Beer advocates this type of controller because "(1) they are arguably the simplest nonlinear, continuous dynamical neural network model; (2) despite their simplicity, they are universal dynamics approximators in the sense that, for any finite interval of time, CTRNNs can approximate the trajectories of any smooth dynamical system on a compact subset of $\mathbb{R}^n$ arbitrarily well" (Beer, 1995, p. 2f). Furthermore, they are very suitable for evolutionary approaches because of their interesting convergence properties - even very small networks can exhibit multi-stable, oscillatory or chaotic behaviour (Beer, 1995, 2006).

The network structure I employ in most models is a partially layered control network in which a layer of input neurons projects onto a layer of fully connected inter-neurons which, again, projects onto a layer of output neurons. However, in individual models this structure is modified, as indicated locally.

The dynamics of neurons in a CTRNN is governed by

$$\tau_i \frac{da_i(t)}{dt} = -a_i(t) + \sum_{j=1}^{N} c_{ij} w_{ij} \sigma(a_j(t) + \theta_j) + I_i(t) \tag{3.2}$$

where $\sigma(x) = 1/(1+e^{-x})$ is the standard sigmoidal function, $a_i(t)$ the activation of unit i at time t, θ_i a bias term, τ_i the activity decay constant and w_{ij} the strength of a connection from unit j to unit i. The $n \times n$ connectivity matrix C specifies the existence of synaptic connections between neurons. In cases where the network structure is fixed, $c_{ij} = 0$ for input neurons n_i, $c_{ij} = 0$ for output neurons n_i if n_j is an output or input neuron and $c_{ij} = 1$ otherwise. In some models, I chose, however, to also evolve some of the network structure including the connectivity matrix C.

The biological analogy of CTRNNs frequently adopted is that a_i represents the membrane potential, τ the membrane time constant, θ the resting potential, $\sigma(x)$ the firing rate, w_{ij} the strength of synaptic connections between neurons and I_i network external inputs impacting on membrane potential. As stated above, this biological interpretation of CTRNN dynamics is not relevant to my modelling approach. CTRNNs represent neural dynamics in a more abstract sense, because they can link sensation and motion quickly and transform patterns of stimulation non-linearly over time, thereby maintaining and building interesting dynamical structures.

CTRNNs are actually continuous dynamical systems, but, as stated before, they are simulated using the Euler method (equation 3.1). Applying the Euler method to the above equation (3.2), the following approximation yields:

$$a_i(t+h) = a_i(t) + \frac{h}{\tau_i}(-a_i(t) + \sum_{j=1}^{N} w_{ij} \sigma(a_j(t) + \theta_j) + I_i(t)) \tag{3.3}$$

In order for this equation to approximate CTRNN dynamics sufficiently closely, the τ_i have to be sufficiently large compared to the time-step h (in most models, $h = 1$). In the models here presented, the minimal ration $\frac{h}{\tau}$ set as parameter boundary is 10 but in most models, it is larger than that.

Simulation

CTRNNs are used to model the internal dynamics of the evolved agents controllers. The emphasis of ER is, however, on the *closed loop* modelling, i.e., a whole situation is modelled, not just input output mappings or decoupled neural dynamics. In a diagram that Beer frequently employs to illustrate this idea (figure 3.4), the CTRNN dynamics can be seen as the dynamics in the innermost box (NS). In order to implement the external closure of the sensorimotor loop, i.e., how an agent's actions in the world impact dynamically on its sensations, the body (middle box) and the environment (outermost box) have to be modelled as well.

In the ER models presented in the current dissertation, agent bodies manifest simply as functions transforming particular environmental variables in neural inputs and neural outputs into velocity or force vectors (e.g., wheel velocity, angular joint velocity, directional velocity, ...). Usually, apart from the control network parameters, the GA evolves a sensory gain S_G and a motor

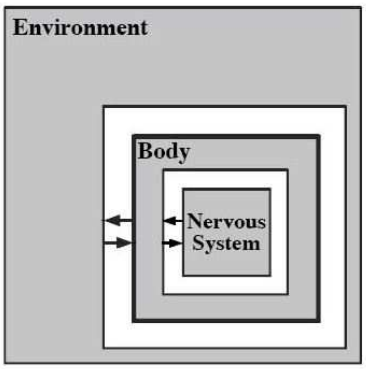

Figure 3.4: Illustration of brain-body-environment interaction, taken from (Beer, 2003).

gain M_G to scale inputs and outputs, which specify the magnitude with which agent and environment impact on each other. The simple transformation functions are coded as part of the class that models the individual.

The simulation classes model the outermost box in Beer's diagram (figure 3.4), i.e., they model a virtual space of some kind in which the state and location of agents and possible external objects are stored and updated, interpreting force and velocity vectors resulting from previous world states and CTRNN outputs. The same time scale is used for neural and environmental dynamics and they are updated alternately with the same time step.

In several models (chapters 6, 7 and 10), sensory delays d have been used. This means that the sensory neural inputs are stored in a queue of length d and fed into the network taking from the front of the queue.

Genetic Algorithm

A genetic algorithm (GA, Holland, 1975) is an optimisation search algorithm for a parameter configuration that performs a heuristic search on the parameter space inspired by the Darwininian principles of heredity, mutation and natural selection that is similar to hill climbing search (but more random).

The search algorithm used in this thesis is a simple generational GA. This means that for a fixed number of generations (typically one or several thousands), a set p of individuals $|p|$ ($|p| = 30$ in this thesis) is used to generate a new generation of equal size and then fully replaced. For each individual $i \in p$, a parent is selected with uniform probability from the $1/3$ best individuals from the previous generation according to the fitness measure F_i (i.e., truncation selection). I implemented non-sexual reproduction, i.e., an individual's genotype is a mutated clone of the single parent's genotype. I use real-valued genes $\in [0,1]$ and vector mutation (e.g., Beer, 1996) as mutational operator. This means that the genotype is mutated by adding a random vector of magnitude r (magnitude Poisson distributed) in the n-dimensional genotype space to the genome. If mutation of a gene exceeds the gene boundary, it is *reflected*, i.e., the amount by which the gene boundary is exceeded is subtracted from the gene boundary to yield the new gene value.

Genes are interpreted as network parameters τ_i, θ_i and w_{ij} and S_G and M_G. The parameter ranges vary between simulations and are specified locally. Typically, $w_{ij} \in [-8,8]$, $\theta_i \in [-3,3]$, and these values are mapped linearly to the specified target range. The minimal value for τ_i is

ca. $20h$ and the maximum value for τ_i is in the order of magnitude of the duration of a trial or a meaningful action in the task. M_G, S_G and τ_i are mapped exponentially to their target ranges, which means that the inter-individual differences between that the GA works on are more fine grained for small values of M_G, S_G and τ_i than for large values of M_G, S_G and τ_i. In some cases, I also modelled the network structure, i.e., I interpreted genes using step functions to determine the existence of synaptic connections c_{ij} or, in some cases, for the existence of inter-neurons n_i.

Typically, fitness evaluation is computed from several evaluation runs. In the models here presented, I either averaged fitness from several trials or I used an exponentially weighted fitness average such that for n evaluations

$$F(i) = \sum_{j=1}^{n} \left(F_j(i) \cdot 2^{-(j-1)} \cdot \frac{1}{2^{-(j-1)}} \right) \tag{3.4}$$

where $F_j(i)$ gives the fitness on the j^{th} worst evaluation trial for individual i. This evaluation technique gives more weight to worse evaluations and thereby rewards the generalisation capacity of the evolved networks. This means that it helps to avoid that evolutionary search gets stuck in a locally optimal trivial solution that stably yields a high score for some parameters of the task. At the same time, it rewards the evolution of such locally optimal behaviour as compared to no sensible behaviour at all, by still giving some fitness for solving parts of the problem.

3.3.2 Simulation Models as Scientific Tools

After explaining what ER simulations are and specifying the technical details of the ER simulation models conducted in the research presented in the current dissertation, I now explain what their contribution to science consists in. This subsection is a very important part of this dissertation, as the common denominator of the models presented in this thesis is their methodological value for a science of human behaviour, cognition and mind.

ALife and ER simulation models are different from computer or simulation models in other scientific disciplines, such as theoretical physics, biology or sociology. The function of such models is, typically, to fit and describe an empirically gathered data set, thereby generalising its structural properties and predicting future measurements. ALife modelling is a more *generative* modelling approach. In clarifying this assertion, I reproduce some of the arguments and positions presented in (Rohde & Stewart, 2008; Di Paolo, Rohde, & Iizuka, 2008b; Beer, 1996; Di Paolo, Noble, & Bullock, 2000; Harvey, Di Paolo, Wood, Quinn, & Tuci, 2005).

Di Paolo et al. (2000) argue that ALife simulation models are to be understood as 'opaque thought experiments':

> "it is reasonable to understand the use of computer simulations as a kind of thought experimentation: by using the relationships between patterns in the simulation to explore the relationships between the theoretical terms corresponding to analogous natural patterns" (Di Paolo et al., 2000).

Simulation models are guaranteed to only produce phenomena that logically result from the premises built into the model as there are no possibly interfering external variables as in complex real-world science. Thereby, they can generate proofs of concept of the kind of processes that can produce a certain kind of phenomenon under certain circumstances - or not. Importantly, through

the use of digital computer technology, simulation models can go beyond human cognitive limits or prejudices. How dynamical systems, in particular non-linear dynamical systems evolve in time is extremely difficult to grasp and intuit without the help of computer simulations.

A good example is Hinton and Nowlan's (1987) simulation model of the Baldwin effect in evolutionary biology: the mechanism had been proposed but not credited because, at first glance, it appeared to propose Lamarckianism. Only with the help of a simluation model, it could be established beyond doubt that lifetime adaptation can aid the evolution of biological traits within a Darwinian framework. "A proposed mechanism that had not been perceived as convincing because it was counterintuitive and difficult to understand had been made credible with the help of a computational model" (Rohde & Stewart, 2008). As a result of this conceptual contribution, the Baldwin effect has become a widely acknowledged concept in evolutionary theory.

This power of simulation models to counter our intuitions and go beyond our imagination, at the same time, makes them more difficult to work with than 'armchair' thought experiments. This is where the 'opacity' comes in: "Due to their explanatory opacity, computer simulations must be observed and systematically explored before they are understood" (Di Paolo et al., 2000). After producing a simulation result, a 'pseudo-empirical' investigation of the simulation follows, in order to understand and explain how exactly it works. Different variables are monitored over time and parameters and conditions are modified in order to discover the systematicities governing the simulation. Such exploration is, in a way, similar to hands-on scientific work, but has the benefit that the *explanandum* is fully controllable, simpler, fully accessible and experiments are easily reproducible. Therefore, it is easier to derive general principles and formal rules governing the simulation dynamics, insights that can then be fed back into the original scientific community to inform theory building.

Harvey et al. (2005) elaborate on the scientific function of ER simulation models in Cognitive Science, using examples from ER simulation research on homeostatic adaptation, the origins of learning and development. As important features of ER simulations, they identify the *minimisation of complexity and prior modelling assumptions*. In the light of the frequent criticism of ALife modelling that it is difficult to conceive how it would scale up (e.g., Kirsh, 1991), it may seem surprising that minimalism is perceived as a merit. Many AI modelling approaches aim at approximating human or real brain complexity as closely as possible (e.g., Markram, 2006). The problem with this kind of approach is that quickly the model becomes as opaque as the original phenomenon, whilst not generating useful generalisations or abstractions.

One of the most passionate proponents of a minimal modelling approach is Beer (1996). When dealing with complex dynamics, even systems that seem very simple at first glance can generate surprisingly complex behaviour. (e.g., Beer, 2003, 1995). Therefore, dynamical principles should first be properly analysed and understood in the most simple and abstracted case, to get intuitions about the kind of dynamical phenomena that exist in sensorimotor interaction, develop tools to study them and then build up complexity gradually. He talks about minimal simulation models as 'frictionless brains' in analogy to Galileo's 'frictionless planes' (personal communication 2005) that allow us to do the mental gymnastics to build intuitions, form concepts and hypotheses in order to ultimately advance with real world scientific work and explanation.

ALife simulation modelling is different from and goes beyond formal description and fitting

of an empirically gathered data set because its results are more conceptual and abstract than quantitative predictions and impact on theory building as well as the scientific practices of designing experiments and interpreting data. An important point to note is that the generative modelling that this section emphasised does in no way contradict, exclude or oppose the possibility of descriptive data-driven modelling. We identify descriptive and generative modelling in psychology as "two poles [...] [that] define a continuum of dynamical approaches" (Di Paolo et al., 2008b).

As concerns the models presented in this dissertation, they can be seen as case studies that emphasise different scientific functions of ER simulation models. The models of synergies (chapter 4) and of value system architectures (chapter 5) are predominantly generative models in the 'opaque thought experiment' sense outlined above. They strongly idealise the original phenomenon observed. The model of synergies (chapter 4) models sensorimotor behaviour as it is observed and mainly cashes out the capacity of simulation models to exceed our cognitive grasp of non-linear dynamics, in order to verify theoretical concepts, generate new hypotheses and suggest further experiments to empirical researchers and synthetic experiments to modellers. In so far, it targets scientific practice. The model of value system architectures (chapter 5), on the other hand, exploits pre-dominantly the fact that simulation models can take us beyond our intuitions, illustrate inconsistencies in conceptual arguments and point out implicitly held prior assumptions, which is more relevant to philosophical debate and theory building than to hands-on experimental practice. The models of perceptual crossing (chapter 6 and 7) and adaptation to sensory delays (chapter 10), are a bit closer to the descriptive pole because the experimental work modelled follows a similar minimalist agenda which allows stronger analogies (see section 3.4 below). Even though they also generate proofs of concept and counterintuitive insights, some direct and quantifiable predictions or measures for gathered data and future experiments results from these models. This use of ER simulation models tries to get the best of both worlds by generating concrete predictions and hypothesis for further experimentation as well as to contribute to the philosophical debate which surrounds the perception research they model, as argued in section 3.6 below.

3.4 Perceptual Supplementation and Minimal Experimental Approaches

In this section, I introduce a minimal experimental approach that is based on research in 'Sensory Substitution' (Bach-y Rita et al., 2003) and that is in many ways ideologically related to the minimal ER simulation approach described in the previous section 3.3.2. This approach has been proposed and advocated by the *Groupe Suppléance Perceptive* (GSP) of the Technological University of Compiègne under the name of 'Perceptual Supplementation' (Lenay, Gapenne, Hanneton, Marque, & Genouëlle, 2003). The models presented in chapters 6, 7 and 10 model results from this strand of minimalist experimental research and chapters 9 and 11 present results from my own experimental work within the mentioned group using the techniques here described. Before outlining the GSP's minimalist variant of Sensory Substitution technology, I briefly summarise Bach-y-Rita's pioneering work. I then relate this approach to other work in sensorimotor plasticity.

In 1963, Bach-y-Rita et al. have started a research program of building prosthetic devices for blind people that allow to substitute for aspects of their visual sense for tactile signals representing visual information (Tactile Visual Sensory Substitution, TVSS; e.g., Bach-y Rita, Collins, Saud-

ers, White, & Scadden, 1969; Bach-y Rita et al., 2003). Equipped with a head-mounted camera that relays pixeled images to arrays of tactile stimulators (on the belly, the fingertip, the back, the tongue,...), congenitally blind people can be trained to perform tasks that are normally considered visual tasks, such as face recognition, catching a ball (which requires 'hand-eye-coordination'), or recognising shapes. Bach-y-Rita sees this technology as a direct extension of the principle of a blind person's cane: even though the cane produces tactile stimulation of the palm of the hand, blind people use it to perceive objects at a distance. As they get used to navigating with a cane, the automated swaying movements and the vibrations in the palm of the hand that holds the cane disappear from their conscious experience and, instead, blind people perceive external objects, such as steps, doors, puddles, etc. In a similar way, when trained with the TVSS, subjects employ visual language to express their experiences, and optical illusions have been reproduced in subjects trained with the TVSS (Bach-y Rita et al., 2003). This fascinating research program, which over the years has been applied also to other sensory disabilities (most noticeably, equilibrial disabilities) continues vividly despite Prof. Bach-y-Rita's recent lamentable death, in his own department and in other groups, who have taken up the idea and built similar devices, exploring also other sensory channels, such as the auditory to visual Sensory Substitution in the vOICe system (Amedi, Stern, Camprodon, Bermpohl, Merabet, Rotman, Hemond, Meijer, & Pascual-Leone, 2007), showing that the principles of this kind of sensorimotor adaptation hold more generally. The term 'Sensory Substitution' has become the general term for technology that records signals associated with one sensory modality and, through the use of technology, transforms it to stimulate, non-invasively, sensors of another sensory modality (Lenay et al., 2003). The GSP is one of the research groups who take part in this effort to build prosthetic devices relying on Sensory Substitution technology.

Apart from its practical prosthetic use to improve the lives of people with sensory disabilities, the fact that this technology works the way it works teaches us some interesting lessons about the nature and sensorimotor origins of human experience and perception. As Hurley and Noë remark, in TVSS "the qualitative expression of somatosensory cortex after adaptation appears to change intermodally, to take on aspects of the visual character of normal qualitative expressions of visual cortex" (Hurley & Noë, 2003). This fact seems difficult to reconcile with the reductionist ideas of functionally dedicated brain areas whose activation is the physical correlate of experiences of a certain modal quality. It thus gives evidence for their "dynamical sensorimotor hypothesis" according to which "changes in qualitative expression are to be explained not just in terms of the properties of sensory inputs and of the brain region that receives them, but in terms of dynamic patterns of interdependence between sensory stimulation and embodied activity" (Hurley & Noë, 2003).

While I wholeheartedly agree with the second part of their argument (i.e., that changes in qualitative experience are to be explained as well in terms of dynamical patterns of sensorimotor interdependence), I am not 100% comfortable with the first part of their argument, i.e., that there is an intermodal transfer of experience and that information received by tactile sensors has visual qualities. This way of thinking bears some remainders of a cognitivist world view in that it presumes experience to come in one of five (or so) pre-defined modal flavours and that these get swapped over when training with Sensory Substitution devices. If you take sensorimotor theories

of perception seriously, you have to rid yourself of obsession with sensory channels, or, as Barbara Webb put it at the Neuro-IT summer school in Venice in 2006, you can only sensibly assert that there are three senses (chemical, mechanical and thermal) or otherwise, you have to accept that there are infinitely many senses (conversation overheard, I added the specification in brackets). This is not to deny that certain classes of experiential qualities are associated with certain classes of perceptual activity or certain sensors. It is just the application of the idea that outside the cognitivist premise, no *a priori* link between the mechanical level (types of receptors) and the functional/meaning level (infinitely many senses, such as sense of colour, direction, shape, posture, ...) can be presumed. In an enactive or non-computationalist view, differences in quality are part of the *explanandum* and should thus not be evoked, without justification to form part of the *explanans*, and particularly not in explaining experiences that do not stem from our natural senses.

This observation resonates with a related observation by Lenay et al. (2003), who criticise the term 'Sensory Substitution' for the described technology as "misleading and in many ways unfortunate" (Lenay et al., 2003). Under close conceptual scrutiny, it becomes clear a) that what people with sensory disabilities gain from this technology are not senses (i.e., receptors), but new perceptual qualities and b) that there is no substitution of the absent sense but rather an augmentation or supplementation of the perceptual world. Thus, what can be observed is much more interesting than simple substitution of missing sensors. Such 'real' sensory substitution (e.g., cochlear or retinal implants) have received much less attention in Cognitive Science literature because they lack the following characteristic:

> "These tools [Sensory Substitution Devices] make it possible to follow with precision the constitution of a new sensory modality in the adult. In particular, by providing the means to observe and reproduce the genesis of intentionality, i.e., consciousness of something as external (the 'appearance' of a phenomenon in a spatial perceptive field), these tools make it possible to conduct experimental studies in an area usually restricted to philosophical speculation" (Lenay et al., 2003).

Lenay et al. propose, therefore, to use the term 'Perceptual Supplementation' (*Suppléance Perceptive*) rather than 'Sensory Substitution'. Bach-y-Rita acknowledges a similar conceptual limitation of the term when remarking that the applications for this technology are open-ended and "could be considered to be a form of sensory augmentation (i.e., addition of information to an existing sensory channel)" (Bach-y Rita et al., 2003) rather than just a substitution, a proposal that explicitly underlies Nagel et al.'s research (Nagel, Carl, Kringe, Märtin, & König, 2005) on human adaptation to an artificial compass sense.

In my own experience of discussing this technology with researchers from different disciplines I have found several times that, even if they are generally sympathetic towards enactive, dynamical and sensorimotor theories, they are misled by the term 'Sensory Substitution' or its interpretation in Cognitive Science literature. Prinz's (2006) critical response to Noë's book 'Action in Perception' exemplifies this unfortunate misunderstanding: Prinz writes that in order for TVSS systems to provide evidence for enactive theories of perception it must be shown that "experience of using the apparatus is like vision, and [...] that it takes on this visual quality in virtue of the fact that subject learn to associate its inputs with the kinds of motor responses that are usually reserved for vision" (Prinz, 2006). Prinz accepts evidence for the latter condition but "seriously doubt[s]

that these subjects experience anything visual" (Prinz, 2006), pointing out that experience of distal objects through tactile sensors forms part of our natural perceptual experience already, such as "when we tap an object with a cane we feel its shape and texture; when we drive, we feel the surface of the road" (Prinz, 2006). Prinz' observations are fully in line with the positions argued by Bach-y-Rita et al. (2003) and Lenay et al. (2003), who explicitly draw the connection between the technology they employ and more rudimentary devices such as a blind person's cane. I have myself nothing to add or object to his analysis, other than that it does not "[put] the Brakes on Enactive Perception" (title of Prinz, 2006) but much rather puts the brakes on the slightly misleading interpretation of Sensory Substitution technology that Noë seems to provide (I have not read his book myself); an interpretation that is suggested by the misleading label 'Sensory Substitution'. In my understanding of the enactive approach, it does not entail that the experiences afforded by Sensory Substitution devices are in any fundamental or ontological way different from experiences produced by the skilled use of other technological devices. Much rather, what makes this technology special, in an instrumental way, is its great potential for the controlled scientific investigation of how mastery of sensorimotor contingencies produces qualitative experience.

Despite my unhappiness about the term 'Sensory Substitution', I sometimes refer to the described technology with this label because the term has established itself. In this thesis and in other contexts where it is known, I prefer to call it 'Perceptual Supplementation' (PS). I argue below why I have come to increasingly refer to the experimental aspects of my work with rather bulky labels, such as 'minimalist experiments on human sensorimotor behaviour' or 'experiments on minimal sensorimotor adaptation'.

The GSP have identified the potential of investigating "minimal forms of sensori-motor coupling" (Lenay et al., 2003) afforded by Perceptual Supplementation technology to study the fundamentals of how aspects of our perceptual experience are constructed rulefully upon familiarization with new sensorimotor couplings. Lenay's habilitation 'Ignorance et suppléance : la question de l'espace' (Lenay, 2003) presents results from a series of experiments using minimal Perceptual Supplementation technology to investigate the fundamental basis of spatial experience. The approach the group has taken in investigating this question has been very similar to the minimalist program in ER simulation modelling described in section 3.3 and which Inman Harvey (personal communication) once suitably characterised (referring to minimalist simulation modelling) as to "throw as much bath water out as possible, whilst keeping the smallest possible baby": simplifying Perceptual Supplementation technology to the extreme (one photo-receptor attached to the finger of a participant's hand that produces a bit sequence of on-off tactile signals), the group have identified the minimal condition under which the interesting changes of experience upon familiarisation occur: for the case of exteriorisation of a perceived stimulus (i.e., an object at a distance), a minimal movement space of two joints and continuous swaying movements as strategy have been identified to lead to the perception of a stimulus as distant and 'out there', whilst one-jointed movement or lateral displacement of the receptor evoke the sensation of touch. The sensorimotor contingency rules that underlie this perception have been mapped out and analysed, and these fundamental findings form the basis for incrementally more complex further experiments, building up on the tractable, controllable and analysable experimental findings previously recorded. The experiments on perceptual crossing by the same group that are described and modelled in chapter

6 and 7 follow a similar minimalist agenda, starting from the simplest scenario possible (one-dimensional environment (Auvray, Lenay, & Stewart, 2008)) and incrementally complexifying the experimental set-up (two-dimensional world, unpublished work) to identify differences and similarities and explain them with reference to the minimal differences in sensorimotor couplings. In many ways, the research on the fundamentals of space perception is the main source of inspiration for the experiments on adaptation to sensory delays and changes in experienced simultaneity described in the second part of the current dissertation (chapters 8 -12).

Recognising that PS technology is not in any fundamental way different from the skilled use of technological objects, such as a car, a cane, puts this research at one pole of a continuum of related research on perceptual learning, psychophysics or sensorimotor adaptation. In this sense, more traditional research on sensorimotor disruption and adaptation, such as experiments with prism goggles (Kohler, 1962; Welch, 1978) are, in my perception, not in any fundamental way different in their potential for the scientific explanation of the embodied sensorimotor bases of perception. While I agree with Lenay et al. that PS technology is particularly interesting because it allows the study of the 'constitution of a new sensory modality in the adult', there is a clear continuum in the degree to which the qualitative experience associated with adaptation to new sensorimotor couplings resembles the previously associated experience. I see this difference as a quantitative difference in the potential to which research on either pole can explain perceptual qualities independent of prior knowledge (or *know-how*) and, in particular, to address questions of the origins of sensory modalities.

The kind of minimalist embodied and dynamical program outlined by the GSP is very promising for the scientific study of perceptual experience. I applied this approach in the experimental study presented in chapters 9 and 11 and, as said in the conclusion chapters 12 and 13, I aim to work within this approach in my future research. The conceptual analysis by Lenay et al. (2003), however, does not only hold for research in PS, but also, more generally, for research on sensorimotor adaptation, perception and certain types of psychophysics, which, for my purposes, are equally interesting. As a pleasant side effect, extending the methodological scope helps to avoid the outlined labelling problems concerning 'Sensory Substitution' vs. 'Perceptual Supplementation'. As mentioned above, in lack of a better label that subsumes the areas of perception research that I am interested in, I refer to the experimental dimension of my work with tags such as 'minimalist research on sensorimotor adaptation'.

3.5 The Study of Experience

In this section, I address the difficult methodological issues around studying and explaining experience that become relevant in the later parts of my DPhil; indirectly in the models of perceptual crossing in chapters 6 and 7 and directly in the interdisciplinary study of adaptation to sensory delays and experienced simultaneity in chapters 8-12. Experience is an inherently subjective phenomenon and only accessible to us through our own first person what-it-feels-like. Science, on the other hand, is about observation and measurement from the outside. It uses third person methods and can therefore categorically not be applied *directly* to the study of experience. Therefore, a purely scientific explanation of cognition is doomed to leave out one of its most defining characteristics, i.e., subjective qualities.

In the computationalist paradigm, this problem has been widely dealt with by, more or less, ignoring it as concerns methodological debate, even though it has been prevalent in the philosophy of mind (qualia debate). This reluctance to explicitly deal with the experiential aspect of cognition results from the historical context in which Cognitive Science arose, i.e., as an opposition to Behaviourism that made the use of mentalistic language credible by putting it in a context of scientific rigour that introspectionist psychology was missing (cf. chapter 2 section 2.1). While the aspiration to live up to scientific standards is honourable, it prohibits the study of experience, which is not accessible in its fullness to purely scientific investigation. The Cognitive Science of experience, therefore, finds itself in a schizophrenic situation where it tries to deal with experience whilst pretending not to be dealing with experience.

The neurophenomenological approach developed by Varela (1996) argues how, within the enactive paradigm, first and third person methods can be combined in order to interdisciplinarily tackle problems of experience. Subsection 3.5.1 gives a short outline of phenomenology as a first person method and introduces Varela's argument, concluding that this approach is preferable to approaches that claim to be purely scientific. Section 3.5.2 argues that the approach taken in my research, which relies on the more rudimentary second person methods used in psychophysics to observe and quantify perceptual judgment can be used in a similar spirit, even if, traditionally, this has not been the case.

3.5.1 First and Second Person Methods as Credible Sources of Knowledge

Chalmers (1995) coined the term 'the hard problem' for the paradoxical difficulty that representationalist Cognitive Science has in explaining the existence of experience in the first place: computational theories of mind can describe functional mechanisms that bring about physically measurable results that share certain structural similarities with physically measurable variables in the brain or human behaviour, which again correlate with the occurrence of particular classes of conscious experiences. But, having a functional and mechanistic description of this kind, the question that remains is why should such a functional unit produce experience at all, rather than just to perform its mechanistic function without experience? This problem is also referred to as the 'qualia' problem or 'the explanatory gap' (Levine, 1983).

Physical *correlates* of consciousness can, to a certain degree, be identified, but they do not *causally explain* the occurrence of conscious experience. Within an approach whose explanatory domain is the material and functional, conscious experience appears to be an unnecessary and causally irrelevant extra, an epiphenomenon. Or, if it bears a functional role, this role can be formally described, reproduced and inserted into the model as a new functional module - but this again raises the question of why there should be any experience at all, leading to a *regressus ad infinitum*.

In a response to Chalmers' statement of the hard problem, Varela (1996) proposes his neurophenomenological approach as a remedy. He briefly reviews the existing theories of consciousness, characterising them along four axes (including the prevailing functionalist approaches; the reader is referred to this scale in order to localise the approach taken here in the landscape of existing theories of consciousness). One of the groups, amongst who he counts himself (and amongst who I count myself), are those who acknowledge that subjective first person experience

is irreducible and plays a central role in a theory of consciousness.

Varela reappraises the classical phenomenological approach established by Husserl (e.g., Steiner, 1997, recent edition of Husserl's lifework ca. 1886-1938) during the *fin du siècle* which promotes phenomenological reduction (see below) as a method for the systematic exploration of one's own experiential world. Varela quotes Merleau-Ponty to establish a first intuition about the link between the first person study of experience and the scientific study of cognition:

> "To return to the things themselves is to return to that world which precedes knowledge, of which knowledge always speaks and in relation to which every scientific schematization is an abstract and derivative sign language, as the discipline of geography would be in relation to a forest, a prairie, a river in the countryside we knew beforehand" (Merleau-Ponty, 2002, recent edition; French original published in 1945), cited in (Varela, 1996).

The reason why many Cognitive Scientists are uncomfortable with the phenomenological tradition is that it appears to be a variant of introspectionist psychology, which, through its lack of intersubjective and methodological standards, made it possible for Behaviourism to become powerful and prohibit the scientific consideration of mind (cf. section 2.1).

There are certainly some commonalities between these approaches that rely on the investigation of the mental and its verbal explication. After all, they are both first person approaches. Varela is right, however, to point out that phenomenological reduction as a method is much more credible. Firstly, it explicates the reflexive and reductive aspect of the act of self-observation, accounting for the nature and source of the introspective activity, which introspectionism left implicit.[2] Secondly, by explicitly including methods of communication and description into the approach and acknowledging its reciprocal causal effect of shaping and modifying the experiential world, the results of phenomenological reduction can stabilise in one's own account and ultimately also become subject to social debate and inter-subjective consensus. Thirdly, Varela argues for the power of intuition, not as an erratic mood swing, but as stable common sense beyond logic that informs all aspects of our life, including scientific activity. This powerful role usually goes unacknowledged in objectivist world views and is at the root of scientist chauvinism and the discarding of first person methods. Fourthly, these standards of generating communicable descriptions, stabilising one's own experience and intuition and mastering the reflexive stance do not come naturally but require training and discipline. Phenomenological reduction is not a *scientific* method of reproducible measurements. In the explication of techniques and issues, however, it certainly comes closer to scientific standards than naïve introspection.

I believe that the lack of appreciation of these merits, which, pragmatically, give it a clear competitive edge over naïve introspectionism, but not necessarily an ontologically different status, stems from failure to recognise just how bad naïve introspection performs in comparison. I believe so because I myself went down that garden-path. By this I do not mean that introspection is fallible in the sense that it does not always concur with the 'objective' observer perspective - systematically and stably occurring illusions or misjudgements that bring the first and third person perspective in conflict (e.g., flashbulb memories; cf. Eysenck & Keane, 2000, p. 226f) are as real an experience

[2] Steve Torrance (personal communication) rightly remarked that, in this sense, even the term 'introspection' is misleading: it appears to suggest the observation of an internal as if it was just a shift of focus from observing the external. The self-referential and reflexive nature of introspecting would be much clearer from the term 'autospection'.

as me seeing the screen of my laptop in front of me right now and can be equally informative for understanding mind, if not more. What I refer to is the bad quality of spontaneous subjective reports and the lack of consistency and structure in common introspection. The apparent stability and consistency of our everyday perceptual and experiential world makes us believe that it is not a big deal to observe and report it.

When learning about the research with 'second person methods', i.e., interview techniques to gather experiential data, I first realised how wrong this assumption is, an error that was later painfully confirmed when straight forwardly querying the experimental participants of the study described in chapter 9 about their experience of the task: they were just baffled, shrugged and did not answer anything useful at all. Research on second person methods develops techniques that can, to a certain degree, compensate for the naïvety of individuals untrained in systematic observation and documentation of their experiential world and thus yield useful reports even from naïve subjects (e.g., Petitmengin, 2006; Vermersch, 1994). Petitmengin states the problem as follows:

> "How many of us would be able to precisely describe the rapid succession of mental operations he carries out to memorise a list of names or the content of an article, for example? We do not know how we go about memorising, or for that matter observing, imagining, writing a text, resolving a problem, relating to other people... or even carrying out some very practical action such as making a cup of tea. Generally speaking, we know how to carry out these actions, but we have only a very partial consciousness of how we go about doing them" (Petitmengin, 2006, p. 230).

Petitmengin gives a much more detailed account of the difficulties with untrained reporting of experiential data in the given source. If the reader is in doubt, it will be much easier to become convinced if he or she tries to generate a verbal report of the phenomenology of searching the cited article on the Internet - or just asking any person around them to report theirs. The result will be very poor because untrained introspectors suffer from "unstable attention, absorption in the objective, escape into representation, lack of awareness of the dimensions and level of detail to be observed, impossibility of immediate access" (Petitmengin, 2006, p. 239). Bringing together techniques from different areas, such as phenomenology, Buddhist meditation and research on consciousness taking as a mnemonic technique, Petitmengin has developed an interview technique that she argues leads to reliable and validatable experiential data.[3] The most impressive proof of the effectiveness of this technique is from its application in non-pharmacological epilepsy therapy, where over therapeutic session using her interview techniques Petitmengin trained epileptic patient's to become aware of and describe their experience of the 'aura' state preceding a seizure and could thus improve their seizure anticipation and suppression skills, yielding a therapeutic effect comparable to or better than benchmark pharmacological treatments (Petitmengin, 2005; Le Van Quyen & Petitmengin, 2002).

Having argued that the study of experience by skilled interviewers or skilled phenomenological reducers produces more useful and reliable experiences and experiential reports than just asking your neighbour, how can these results be linked to results from third person science without

[3]The fact that such an interview and its setting also influences and modifies experience is not *a priori* a problem. For an approach that aims at minimising this impact of the second person and come close to 'experience in the wild' see Hurlburt (Hurlburt & Schwitzgebel, 2007).

stepping into a reductionist trap? In order to explicitly link the experiential and physical aspects of cognition and to communicate this link, the experiential has to manifest in some form in the physical world. Experiential reports, as they result form second person techniques, or, if I report my own experience, from first person experiential exploration, are such manifestations that can be treated, to a certain extent, as data. It has, however, to be stressed that, by calling reports a manifestation of the mental in the physical, I do not intend to imply a superiority of the physical manifestations or suggest the possibility to reduce experience to the act of reporting/measuring it. This step just serves as a method for interfacing the two, a possibility of integrating the experiential and the scientific knowledge into an interdisciplinary story that is half and half.

So, what are the links between the experiential and the physical? Varela remarks that "human experience [...] follows fundamental structural principles which, like space, enforces the nature of what is given to us as contents of experience" (Varela, 1996). Physical structures and regularities constrain and shape our experience - our experiences may be subjective but they are by no means arbitrary. We realise just how regular experience is if we study its perturbations, for instance, during development (e.g., Piaget, 1936), through pathological cases (blindsight, hemineglect, Perceptual Supplementation, ...), under sensorimotor perturbations (e.g., Kohler, 1962) or through altered states of consciousness (e.g., Shanon, 2001) - perturbations that are physical events we can observe, measure and explain.

Varela thus proposes a 'neurophenomenological circulation', whose objective he describes as seeking "articulations by mutual constraints between field of phenomena revealed by experience and the correlative field of phenomena established by the cognitive sciences" (Varela, 1996). This means nothing more and nothing less than making the links explicit that relate the experiential and the scientific results. He gives examples from the neuroscientific study of attention, body image, perceptual filling in, emotion, Libet's (2004) work on voluntary action and his own neurophenomenological explanation of present-time consciousness (Varela, 1999).[4] Again, the most impressive and, in my understanding irrefutable demonstration of the power of this approach is to be found in its application to epileptology (Petitmengin, 2005; Le Van Quyen & Petitmengin, 2002): not only do we study how irregularities of neural activity lead to dangerous and painful seizures, we also study how they lead to altered experiences preceding the seizure (bottom-up). Through the skilled and systematic study of these experiences resulting from abnormal neural activity, the experiences can be transformed, which, ultimately, results in the alteration and control of neural activity (top-down).

An issue that is mentioned but, in my opinion, underdeveloped in Varela's account is the fact that presumably purely scientific accounts of consciousness do exactly the same thing, even if they pretend not to: "It makes us forget that so-called third-person, objective accounts are done by a community of concrete people who are embodied in their social and natural worlds as much as first-person accounts" (Varela, 1996). As a leftover from the behaviourist age, talking about experience or attempting its scientific study is an embarrassment, a cosmetic flaw, which is why the most radical followers of scientism prefer to claim experience does not exist (e.g., Churchland & Churchland, 1998). Research that addresses experiential phenomena, such as the study of the neural correlates of consciousness (Metzinger, 2000), however, has to deal with it by necessity -

[4]The latter two are presented in more detail in chapter 8.

something has to correlate, after all.

There is the clear danger that, in order to keep up the illusion to be fully scientific, research on conscious experience does not explicate its methodological commitments in the first person realm and the presumed nature of its link to the physical. Ironically, the misguided aspiration for scientific rigour introduces conceptual gaps in the explanatory framework. "The line of separation between rigor and lack of it, is not to be drawn between first and third accounts, but rather on whether a description is based or not on a clear methodological ground leading to a communal validation and shared knowledge" (Varela, 1996).

3.5.2 Psychophysics as 'Neutral Territory'

In the conclusion of his proposal of neurophenomenology, Varela writes

> "every good student of cognitive science who is also interested in issues at the level of mental experience, must inescapably attain a level of mastery in phenomenological examination in order to work seriously with first-person accounts" (Varela, 1996).

Personally speaking, I clearly come short of this criterion - I never seriously investigated the techniques of phenomenological reduction (or interview techniques), let alone practiced them. How can I claim to study perceptual experience, how can I claim to be an enactivist?

The truth is that, at the time, I just did not act up to the conviction that these kinds of training were what I should be doing. Therefore, in my study of adaptation to delays, the methodological issues around the experiential aspect of it, i.e., experienced simultaneity, are clearly underdeveloped. I decided to focus on the physical aspects of the task instead.

For future research and in order to compensate for this failure of mine, I want to at least theoretically complete the interdisciplinary enactive methodological framework I propose in this dissertation (see following section 3.6) by outlining how I believe the experiential domain can be methodologically incorporated into the kind of minimal experimental and modelling research I conduct. I promote a different set of second person methods in this section that I believe can complement Varela's neurophenomenological approach, in combination with the minimal modelling and experimental approach sketched. The methods I refer to are the classical measures of perceptual judgment used in psychophysics.

The original statement of the psychophysics research program through the publication of *Elemente der Psychophysik* by Fechner (1966, recent edition; German original published in 1860) is, indeed, so similar to Varela's statement of the neurophenomenological approach that I am surprised this link has, to my knowledge, never been made before.

Against the dominant Cartesian currents at his time, Fechner thought of the mental and the physical as two perspectives of the same thing, like the inside and the outside of a circle, or the heliocentric as opposed to the geocentric perspective of the universe. With reference to Descartes' allegory of the mental and the physical as two clocks that are perfectly synchronised, he remarks that the easiest possibility, i.e., that it is actually just one clock, had not been taken into consideration (Fechner, 1966, p. 4; original in 1860). This perspective implies that asking how one realm links to the other (such as by one being reducible to the other) is an ill-posed question.

He also recognises the importance of the observer status of the scientist (cf. section 3.1):

> "What will appear to you as your mind from the internal standpoint, where you yourself are this mind, will, on the other hand, appear from the outside point of view as the material basis of this mind. There is a difference whether one thinks with the brain or examines the brain of a thinking person. These activities appear to be quite different, but the standpoint is quite different too, for here one is an inner, the other an outer point of view" (Fechner, 1966, p. 3; original in 1860).

Applying these ideas to methods of enquiry he remarks

> "The natural sciences employ consistently the external standpoint in their consideration, the humanities the internal. The common opinions of everyday life are based on changes of the standpoints, and natural philosophy on the identity of what appears double from two standpoints. A theory of the relationship of mind and body will have to trace the relationship of the two modes of appearance of a single thing that is a unity" (Fechner, 1966, p. 5; original in 1860).

Fechner describes the goal of psychophysics enquiry to answer questions like: "what things belong together quantitatively and qualitatively, distant and close, in the material and the mental world? What are the laws governing their changes in the same or in the opposite directions?" (Fechner, 1966, p. 8; original in 1860). This formulation has clear parallels in the neurophenomenological approach.

Where the two positions, in my reading, deviate, is in recognising the importance of closed loop dynamical brain-body-environment interactions:[5] Fechner's vision of how an 'internal psychophysics' of brain physiology would help to identify the direct functional correspondents of sensations, whereas the 'external psychophysics' method he develops and applies 'only' investigates correlations that are mediated through bodily states outs him as more of a localist than I am comfortable with. Similarly, the methods outlined by Fechner are very much restricted to linking sensory stimuli ('inputs') to experience and do not allow for the inclusion of actions or motion into the psychophysical story. In this sense, the original formulation of the psychophysics project is in tension with the enactive or neurophenomenological approach and the circular causality between levels and the agent and the environment it emphasis.

Nevertheless, Fechner's painful awareness of how this method and the dualistic language it adopts lends itself to dualistic interpretation, contrary to his own view of the nature of the link between the mental and the physical, his repeated reassurance that the proposed method produces valid results immaterial of metaphysical questions, whilst hoping and believing that this method would ultimately produce results to confirm his views in a remote future, are not without substance. He certainly did in no way encourage homuncular and representationalist interpretations of his approach, like Baird and Noma's statement that the key question of psychophysics was "how does the human being use sensory and cognitive mechanisms to perceive the type and amount of stimulus energy" (Baird & Noma, 1978, p. 2).

The reason why I believe Fechner's psychophysics approach to be relevant is that, like Varela, he puts his methodological commitments in both the physical and the mental realm and how he believes them to relate open on the table. In the experiential realm, psychophysics investigates and measures perceptual *detection, identification, discrimination* and *scaling* (Ehrenstein

[5] Please note that Poincaré was six years old at the time of the publication of the 'Elements of Psychophysics'.

& Ehrenstein, 1999). The techniques for measuring these perceptual judgments have been used and developed for more than a century. As stated above, the reason why we can study cognition interdisciplinarily is that experience follows fundamental structural principles that relate to physical constraints and sensorimotor invariants. In some cases, these constraints are so strong that they lead to reliable, verbally expressible and intra- and intersubjectively stable results without the need of an expert interviewer or experiencer. The methods to explore the experiential domain that Fechner proposes do not go as deep as phenomenological reduction.[6] However, the reason why the psychophysics approach can address questions of perceptual experience is that it deliberately confines itself to experiential phenomena and judgments that are so primitive that they lead to stable results despite the naïvety of the experiencers investigated.

The major advantage of studying these fundamental dimensions of perceptual experience is that the acts of performing, observing and reporting such perceptual judgments form part of everyday human life. Therefore we are, in a sense, all trained experts in these first person techniques. Also, because they form such an essential part of our world, the inclusion of these perceptual judgments into a scientific framework has not faced a lot of controversy: their quantification makes obvious and intuitive sense, without researchers or audience necessarily even being aware of the metaphysical and ontological questions such an approach raises. Similar methods have also been used in newborns ('high amplitude sucking') and animals (e.g., Melchner, Pallas, & Sur, 2000), leading to speculation about their cognitive domains without appearing to cause a lot of uproar because you can always retreat to a behaviourist stance in justifying your research.

The downside of the common-sense-ness of this method is that it can be easily hijacked by representationalists, elimininativists and behaviourists *because* these results can be adopted into their framework, as it already appeared to have happened to Fechner 15 decades ago. Asking and recording perceptual judgments, though, strictly speaking, a second person method, can easily be treated like just another physical variable to be explained, transcending the functional and the mechanistic realm. This is what I mean by calling psychophysics 'neutral territory': it works regardless of ideological commitment, even if the exact way of investigating phenomena using psychophysics methods and the interpretation of results will be contingent on the choice of paradigm. This uncontroversial nature of psychophysics research can also be seen as its strength: results thus generated will not encounter a lot of resistance on political grounds and may thus help to communicate and illustrate results conducted under the enactive paradigm, which will ultimately benefit its establishment and the refutation of the classical view.

Within a cognitivist perspective, the applicability of psychophysics as approach is usually assumed, more or less explicitly, to be restricted to 'low level' automatic sensory pre-processing that builds up mental representation for cognition to operate on. For some of the most intriguing phenomena studied in psychophysics, such as binocular rivalry or bistable figures (Breese, 1909; O'Shea, 2004), it is not clear in how far they fall into the realm of psychophysics or psychology. Within the enactive view, it is not clear if and how a line can be drawn. It is important to realise that in the original proposal of psychophysics, this restriction had not been allowed for either. When I propose psychophysics as a complementary approach to neurophenomenology, I refer to its original proposal of establishing the link between measurements of perceptual judgments

[6] Please note that Husserl was one year old at the time of the publication of the 'Elements of Psychophysics'.

laid out in the psychophysics method and directly scientifically observable physical phenomena, irrespective of their nature. Advances in technology and mathematics allow the extension of the third person methods associated with psychophysics not just to neurophysiology, but also make its incorporation into more situated and dynamical research programs possible. Rodgriguez et al.'s (Rodriguez et al., 1999) work on neural synchrony and shape recognition, as well as Libet's (2004) neuroscientific study of volitional action, both lines of research that Varela (1996) mentions in his statement of the neurophenomenological approach are very close to what Fechner imagined as 'internal psychophysics'. Similarly, I consider O'Regan et al.'s (O'Regan, Rensink, & Clark, 1999) research on change blindness as in line with the program here outlined.

This is obviously not to argue *against* the neurophenomenological approach - the psychophysics approach can surely not be taken towards all dimensions of experience. However, investigating such rudimentary dimensions of experience can complement the results from classical phenomenology. Both methods have their merits and demerits - psychophysics does not go as deep as phenomenological reduction does. However, perceptual judgments work for everyone and do not involve a transformation of experience through the act of observing it that would go beyond the kind of transformations such acts of self-observation induce on a daily basis.

Besides phenomonology and the mentioned related interview techniques, there is another set of first person and second person methods that produce stable and reliable results and that have been incorporated, explicated and rigorously tested within a scientific tradition over the last 150 years. The primitive experiential aspects these techniques investigate does not require extraordinary reflective or communicative skills, which is both their merit and their demerit, as argued above. It cannot hurt for a good student of Cognitive Science to learn the technique of phenomenological reduction but I do maintain - and this is my only point of real disagreement between Varela's and my view to date - that there are ways of pursuing a good enactive Cognitive Science that takes experience seriously but does not require mastery of phenomenological reduction. It can do so by clearly delimiting itself to certain rudimentary forms of perceptual experience.

3.6 Combining Experimental, Experiential and Modelling Approaches

Having introduced the empirical, synthetic and subjective methods individually, it seems quite clear how they would work together as an alternative interdisciplinary framework. In this section, I want to address the links between these different approaches, in order to elucidate three issues: firstly, the classical reductionist and the non-reductionist enactive approach are juxtaposed. Secondly, the status of simulation modelling in the enactive paradigm is discussed. Thirdly, I argue that this approach is truly interdisciplinary, rather than just multidisciplinary. These ideas have been presented in part in (Rohde & Di Paolo, 2006a; Di Paolo et al., 2008b, 2008a).

In a simplified view on the computationalist paradigm, AI modelling forms the intellectual centre-piece of a reductionist program (see figure 3.5 (A)): philosophy establishes the relation between 'qualia' and neural states, which ultimately results in a reduction of the mental to the physical based on functional causal role. This reduction is via a formal AI model which captures the essence of brain functionality and which, in principle, could be variably instantiated; its 'wetware' basis, studied by neuroscientists, is just the way cognition happens to be implemented in nature. In this reductionist view, scientists can quite happily work within each of these levels,

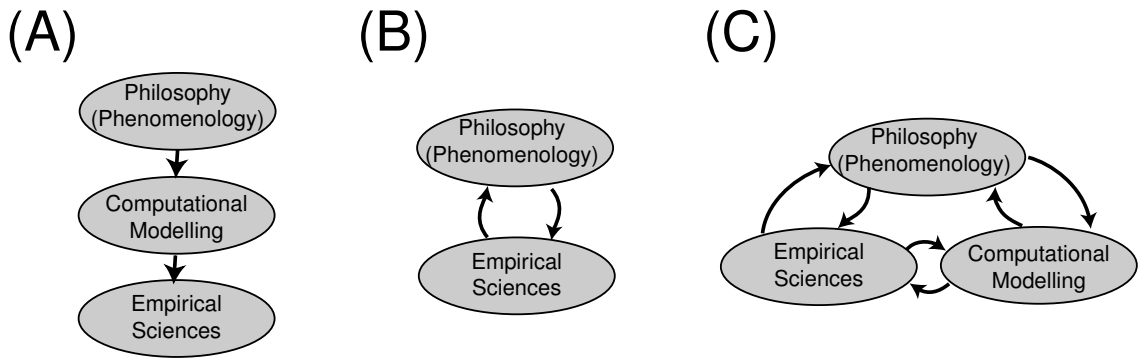

Figure 3.5: Illustration of interdisciplinarity in (A) computationalism, in (B) the neurophenomenological approach and in (C) the interdisciplinary enactive approach proposed.

only occasionally making reference to findings from levels below and above: ultimately, the functional/behavioural level does not depend on its implementation or the mental states it produces.[7] In that sense, this approach is multidisciplinary, rather than interdisciplinary.

The enactive paradigm as a paradigm of non-reductive naturalism, does not have an intellectual centre-piece: as argued in the previous section 3.5, first/second person methods and third person methods are in an active and circular multilogue, exemplified in the neurophenomenological approach (see figure 3.5 (B)) and thereby truly spans levels of explanation, integrating them and requiring proper interdisciplinary activity.

Stewart identifies as one of the two basic requirements for a paradigm in Cognitive Science (besides resolving the mind-body problem) that "it must provide for a genuine core articulation between a multiplicity of disciplines, at the very least between psychology, linguistics and neuroscience" (Stewart, 2008). What is remarkable about this list is that synthetic methods or computer science, from having formed the intellectual centre-piece in the computationalist approach appear to have dropped out of the list altogether. Apart from promoting the enactive paradigm against the prevailing computational paradigm in Cognitive Science, reaffirming the place of computer modelling within the enactive approach to cognition, not as the centre-piece, but as an equal contributor, is one of the core objectives of the current dissertation (figure 3.5 (C)).

In Varela's early work, simulation modelling in the spirit outlined above (section 3.3.2) formed an essential component, as most noticeably reflected in the computational model of basic autopoiesis (Varela, Maturana, & Uribe, 1974). From an initial enchantment with the ALife paradigm in AI, which (at least in some variants) is ideologically so close to the enactive approach (cf. chapter 2), enthusiasm in the enactivist community appears to have cooled down significantly over the decades. The more recent formulation (Varela, 1996) and application (e.g., Le Van Quyen & Petitmengin, 2002; Rodriguez et al., 1999) of the neurophenomenological approach (cf. section 3.5) does not make explicit mention of computational or simulation methods at all.

Part of the responsibility for this trend is probably to be found in the ALife community, which,

[7] A running gag of the students on my Cognitive Science course was the repetition of a phrase we were given in the inaugural lecture: "Cognitive Science is more than just a cocktail of psychology, linguistics, computer science, philosophy and neuroscience" whenever we wanted to point to the fact that in the reality of our studies, apart from some few noticeable exceptions, it was just that: a cocktail.

in my opinion, has increasingly focused on itself and not sought association with the enactive approach - with some noticeable exceptions here at Sussex and elsewhere, (e.g., Beer, 2004; Di Paolo, 2003, 2005) and created a methodological bubble in which interdisciplinary links are, if it all, mainly sought with branches of chemistry, ethology and biology that do not associate themselves directly with the enactive approach or the study of cognition, even though autopoiesis theory was originally one of its main inspirations. The evident explanatory power of simulation models (cf. section 3.3.2) also has led to the re-integration of these techniques in an enactive Cognitive Science outside the ALife paradigm (e.g., Stewart & Gapenne, 2004) but it is undeniable that computer science is a marginalised discipline in the current enactive Cognitive Science.[8]

I propose to bring the generative ER modelling approach to back into Enactivism. In particular, I propose to pair it up with the equally minimalist and enactive perception research in PS. This has a number of crucial advantages. Firstly, by virtue of using similar virtual environments as those employed in ER simulation, no strong abstraction of the behaviour modelled has to be undertaken in order to follow the minimalist agenda described above. Picking a suitable empirical approach (such as PS), ER simulations can be both generative models and descriptive models in the more traditional sense of computational modelling. This means that they can generate quantitative predictions about data already gathered or expected from further experiments.

It is important to point out that the generation of hypotheses from ER modelling makes this such *post hoc* data analysis more credible than an exhaustive exploration of the space of descriptive variables through data mining/data dredging, as, e.g., described in (Ioannidis, 2005). Statistical data mining tools search exhaustively for statistically significant differences in existing data that match the expected pattern. Thereby, such tools increase the probability that differences reported are only stochastic accidents and do not result from underlying structural differences, which is why *post hoc* data analysis is sometimes frowned upon. ER simulation models, in contrast are not directly driven by the data recorded but by functional constraints on the task domain, which means that such accidental stochastic effects do not occur. They generate a small number of very specific predictions about structure in already gathered data that are based on functional and dynamical characteristics of the behaviour modelled, not on statistical patterns in the data modelled. Therefore, predictions generated from ER simulation models of existing data can, in the general case, be treated as if they resulted from a separate study, even if, in any particular case, further experiments or additional control conditions may be required to confirm new descriptive/theoretical insights thus obtained.

The second key advantage is associated with the possibilities of PS research as a stand-alone method outlined in section 3.4: in studying the sensorimotor basis of perceptual experience, PS involves methodological circulation between empirical and experiential methods, which, in the spirit of Varels's (1996) neurophenomenological approach, can naturalise aspects of perceptual experience. I further propose an explicit commitment to the measures of perceptual judgment used in psychophysics, if the *explanandum* allows, for the reasons given earlier. Adding ER models to the picture, the difficult study of the dynamics of sensorimotor behaviour and contingencies (cf. O'Regan & Noë, 2001) becomes more accessible, more formal and more transparent, due to the explanatory potential of ER models described in section 3.3. This is the interdisciplinary

[8] See Fröse (2007) for a discussion of the role of AI in the enactive approach.

framework I propose, and which is illustrated in figure 3.5 (C).

The research presented in this dissertation builds itself up by illustrating, step by step, the mutual links between the discipline of simulation modelling and first and third person methods in figure 3.5 (C), thereby demonstrating that the common root of ALife and the enactive paradigm has not yet been cut: the results of modelling motor synergies (chapter 4) illustrate the mutual link between simulation modelling and the empirical experimental sciences, where models can generate descriptive concepts and generate hypotheses for further experiments. The model on value system architectures (chapter 5) illustrates how simulation models can work like thought experiments in philosophical and conceptual debate, pointing out implicitly held prior assumptions and go beyond intuition. The models of perceptual crossing in a one-dimensional (chapter 6) and a two-dimensional (chapter 7) environment models PS research that in itself adopts a circular and enactive method (figure 3.5 (B)) and therefore shows how simulation modelling can take part in a properly interdisciplinary multilogue, where all arrows in the diagram in figure 3.5 (C) are active. The second part of this dissertation that presents the study of adaptation to sensory delays and perceived simultaneity (chapters 8-12), finally, puts to work the idea that a Cognitive Scientist, in order to claim to work in an interdisciplinary project, has to work interdisciplinarily herself, not just contribute computer simulation models to enactive researchers working with first and third person methods. This assertion underlying the work in the second part of my dissertation, which has not been properly argued or explicated, is critically assessed and relativised in the conclusion chapter 13.

A practice adopted in the later chapters of this dissertation that implement the combination of ER and PS research (i.e., chapter 6 onwards) is to reserve the terms ‘empirical’ and ‘experiment’ for the real world experiments with humans, while the terms ‘simulated’, ‘synthetic’ and ‘model’ are used to refer to the Evolutionary Robotics simulation of the task and its results.

The methodological conclusion (chapter 13) drawn from the entirety of the research presented in this dissertation is, in the light of the abundance of results from the research, altogether very positive about the potential of ER modelling for the enactive approach in general (section 3.3.2 and the interdisciplinary framework sketched in this section in particular. However, probably unsurprisingly, given the novelty of the proposal, I also identify weaknesses and space for further methodological work. Apart from the lack of clarity about the exact role and method in the first person realm already pointed out in the previous section 3.5, the main point requiring further methodological development is the exact structure and benefit of working with empirical, synthetic and experiential methods at the same time. I conclude that this last point developed in this section is only valid in a weak sense.

Chapter 4

ER Can Generate Proofs of Concept and Hypotheses for Science: Linear Synergies as a Principle in Motor Control

---

'The centipede was happy quite,
Until the toad in fun
Said 'Pray, which leg goes after which?'
Which worked his mind to such a pitch,
He lay distracted in a ditch,
Considering how to run.

(Anonymous)

This chapter presents the results from a simulation model I conducted during the first year of my DPhil that investigates a principle in motor control called 'motor synergy'. The term had been invented by the Russian physiologist and biologist Bernstein (1967, recent edition; Russian original published in 1935) for systematicities in motion, and he proposes such systematicities as a principle that helps the nervous system deal with redundancy in motor space. The modelling work is directly inspired by experimental physiological work conducted by Gotlieb et al. in Boston and Indiana (Gottlieb, Song, Almeida, Hong, & Corcos, 1997; Zaal, Daigle, Gottlieb, & Thelen, 1999) on human target reaching. The results in this chapter have been published in (Rohde & Di Paolo, 2005). It also builds up on simulation studies I generated for my MSc dissertation (Rohde, 2004). Other than the models presented in the later parts of this dissertation (chapters 6 and 7 on perceptual crossing and chapters 8-12 on simultaneity perception), the model presented in this chapter is a strong abstraction from and idealisation of the original experiment conducted. However, in comparison to the more conceptual or philosophical model on value system architectures presented in the following chapter, the proofs of concept the model presented in this chapter provides are of more immediate applicability to scientific practice. As outlined in chapter 3 section 3.6, it serves as an example of how simulation models can resonate with experimental research in the cognitive and behavioural sciences.

I introduce the theoretical, experimental and modelling background, as well as the research question to be addressed with this model in section 4.1. Section 4.2 introduces the model, which investigates 'linear synergy' (i.e., a linear relation between torques applied to the elbow and shoulder joints) in a two-dimensional and three-dimensional simulated arm. Evolvability is compared

along two different dimensions of model complexity: dimensionality of Euclidean space and dimensionality of motor space (linear synergies). The results are presented in section 4.3 and they show that, while dimensionality reduction through motor synergies increases evolvability in the given task, dimensionality reduction in space decreases evolvability. These seemingly contradictory results on the usefulness of imposing and releasing constraints in the given simulation model are evaluated as to what they show for motor control task and evolvability in general, as well as in the context of the experimental scientific work on human motor control in section 4.4.

4.1 Background: Motor Synergies

In this section, I outline the degree-of-freedom (DoF) problem in motor control as diagnosed by Bernstein's (1967, original in 1935), as well as his proposal of 'motor synergies' as a remedy (subsection 4.1.1). His biomechanical work has been the inspiration for many experimenters and modellers since it reached the English speaking world after the fall of the iron curtain in 1967 and the evidence for the existence of linear synergies in humans and animals is abundant. Subsection 4.1.2 presents two experimental studies that have been the direct inspiration for the model presented in this chapter and outlines the research question the model addresses.

4.1.1 The Degree-of-Freedom Problem and Motor Synergies

The rhyme with which I started this chapter nicely illustrates what Bernstein (1967, original in 1935) called the DoF problem. If the brain is thought of as a homuncular control organ that controls the state of all muscles and actuators centrally and simultaneously, the task it has to solve is very complex. Trying to describe animal or human behaviour in terms of joint kinematics already involves a large number of DoFs (e.g., 7 in moving an arm). This is the level of complexity aspired in main stream humanoid robotics, keeping many engineers and programmers employed full-time.

The problem of controlling joint positions centrally, however, pales in comparison to the control problem of controlling a living human body centrally. Thinking of motor control in terms of individual muscles, or even motor neurons, the number of DoFs to be controlled when moving an arm quickly exceeds four digits (Bernstein, 1967, original in 1935). Also, while the joints used in robotics are usually exclusively sensitive to the motor signal by the robot controller, biological motor control has to be performed in the presence of *context conditioned variability* (Bernstein, 1967, p. 246ff; original in 1935). The effect of a motor command is sensitive to the anatomical, mechanical and physiological context of the interaction of an agent with its environment, e.g., limb positions, passive dynamics or the state of the peripheral nervous system. Last but not least, the human and animal motor system is *redundant* with respect to the outcome of an action: there are infinitely many trajectories to proceed from a position A to a position B, a condition that Hebb has termed 'Motor Equivalence' (Hebb, 1949, p. 153ff) and humans and animals are very apt at compensating for perturbations, lesions or restraints on the motor system by using different effectors in order to perform the same functional behaviour.[1]

[1] A famous example for this is the fact that characteristics of handwriting are preserved even when forcing a subject to write with its left hand, the mouth or the foot (Kandel, Schwartz, & Jessel, 2000, p. 657).

A homuncular view of how the body could be controlled from a central instance, like a puppet, was common at the time, and explaining how a central organ could manage all this complexity at once seemed a big challenge.[2] Bernstein thought that *systematic relations between effectors*, a concept that he called 'motor synergy', was the answer to the DoF problem. The driver of a car can determine the position of both wheels of the car at a time because they are linked. This link imposes a constraint on the possible wheel positions. However, it only rules out useless wheel positions and does not functionally constrain the motion possibilities of the car. In a similar way, he thought mutual constraints in an organism's motor system could serve to build functional subunits, thereby reducing the effective number of DoFs in a motor task in a beneficial way.

Motor synergies are evident in human and animal behaviour, ranging from human directional pointing (as described in the following subsection 4.1.2) to different types of gaits, posture correction during breathing and hand motion in firing a gun (for a summary of findings see the chapters by Turvey, Fitch and Tuller in (Kelso, 1982)). Bernstein's idea of motor synergies also have strongly impacted on theory building and modelling work in Cognitive Science and motor control (e.g., Arbib, 1981; Grossberg & Paine, 2000; Morasso, Mussa Ivaldi, & Ruggiero, 1983; Sporns & Edelman, 1993).

From an enactive perspective on sensorimotor behaviour, the DoF problem, as defined by Bernstein, does not really pose itself because motor control is not thought of as the result of homuncular central planning. Also, this conception is not free from practical and conceptual problems: as argued in chapter 2, homuncular explanations typically pass the explanatory burden down: is explaining the brain as the 'driver of the bodily car' much easier than explaining the whole system in the first place? Also, Weiss and Jeannerod (1998) remark that "the context in which a motor task is executed strongly influences its organization" (Weiss & Jeannerod, 1998, p. 74). This appears to contradict the idea of functional and structural isolation of motor planning (homunculus) and execution (systematic co-activation of DoFs as functional sub-unit or building block).

However, in the light of the mentioned evidence for systematic relations between motor signals in different DoFs, questions about their nature arise: if motor synergies do not serve the purpose to decrease dimensionality for central motor planning, what is their functional role? How do they emerge from the redundant and high-dimensional movement space? How are they maintained?

4.1.2 Directional Pointing

The particular experimental study that inspired the simulation model here presented is a finding on *linear synergies* (i.e., a linear correlation between torques applied to the shoulder and elbow joint) in human directional pointing by Gottlieb et al. (1997). Targets were arranged spherically and equidistant from the starting position in the sagittal plane. Reaching these targets, the dynamic components of muscle torque (gravitational component removed) applied to the joints were scaled linearly with respect to each other. This systematic relationship does not appear to result from the nature of the task, as it does not produce shortest paths or appear to satisfy any other obvious efficiency or performance criterion.

[2] Bernstein used to demonstrate this to his students by asking them to assume the role of a homunculus and control a system he set up from sticks connected such that they had several degrees of freedom.

Zaal et al. (1999) found the same systematic relationship between joint torques in infants even in the pre-reaching period, even though their attempts to grasp an object are unsuccessful. They investigated infants' reaching behaviour at several stages during their motor development, observing linear synergies throughout the stage-wise development of behaviour. Therefore, linear synergies do not appear to be the outcome of a learning process either. Zaal et al. conclude that "If linear synergy is used by the nervous system to reduce the controlled degrees of freedom, it will act as a strong constraint on the complex of possible coordination patterns for arm movement early in life" (Zaal et al., 1999, p. 255).

Another finding that has to be born in mind is that there are behaviours in which humans learn to break linear synergy. For the case of arm movement, for instance, Weiss and Jeannerod's review on grasping and reaching studies observes that sometimes Cartesian space dominates motor organisation, whereas in other cases (such as in Gottlieb et al.'s (1997) study), joint space dominates the organisation of trajectories (cf. Weiss & Jeannerod, 1998). Therefore, linear synergy does not appear to be a mere fixed physiological constraint on possible arm movements either.

As outlined in section 3.3, one of the methodological advantages of ER modelling is that it does not presume a fixed relationship between the mechanical organisation and functional organisation. Previous modelling approaches to motor synergies (e.g., Grossberg & Paine, 2000; Sporns & Edelman, 1993; Morasso et al., 1983) built in synergy function as a movement building block for composition of complex motion. To the contrary, the ER model presented in this chapter aims at investigating the functional role of motor synergies in a minimally biased way to explore their functional role with minimal prior assumptions. Where do synergies come from? Under which circumstances do they arise? Are synergies epiphenomenal to a structured agent environment interaction or implemented in the control architecture of evolved agents?

If linear synergies are beneficial to the organisation of the modelled task, their existence will lead to an improvement in either performance or evolvability and an exploration of this advantage can generate hypotheses about their functional role in human motor control that can be tested in further experiments. The simulation compares a two-dimensional version of the task with a three-dimensional version to investigate the relation between redundancy in DoFs and spatial complexity. Four different kinds of neural controllers are compared, with and without in-built linear synergies (details are specified in the following section 4.2) to investigate their functional role in and resulting from artificial evolution. The findings are in line with Zaal et al. (1999) in suggesting that linear synergy as a built-in constraint benefits an efficient developmental process.

This exploration is also relevant for robotic engineering and the technical side of ER modelling. In order to be minimal, many Evolutionary Robotics experiments typically do not involve high levels of redundancy. The results here presented demonstrate how imposing the *right* constraints along the *right* dimensions can impact on evolvability and the nature of the solutions evolved. For instance, as a consequence of the insights gained here, I successfully evolved a linear synergy controller in a variant of the model of two-dimensional perceptual crossing presented in chapter 7, even if these results are not presented as part of the present dissertation.

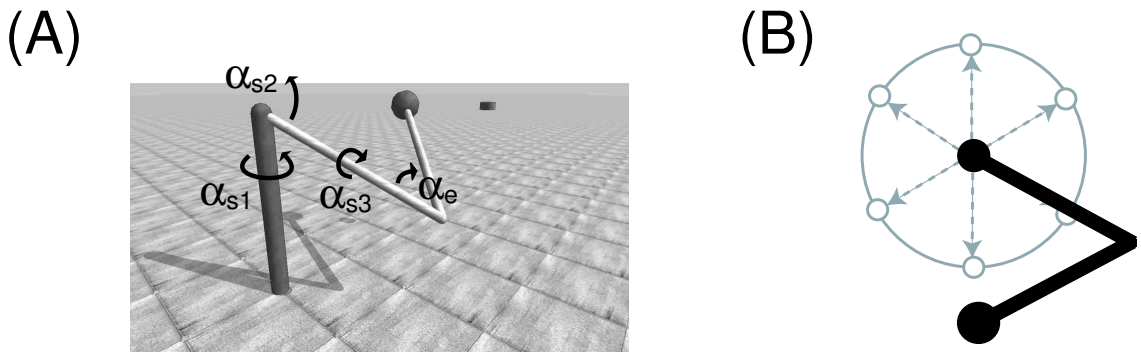

Figure 4.1: (A) Visualisation of the simulated arm. (B) Schematic diagram of the task.

4.2 Model

A robotic arm is evolved to reach to one of six target spots on a horizontal plane. Unlike the other simulations presented in the current dissertation, the model presented in this chapter has been programmed in C++. The reason for this choice is that the simulated arm is implemented using the open source physics simulation library 'Open Dynamics Engine' (ODE, Smith, 2004). However, the algorithmic details of GA, numerical integration and CTRNN control dynamics described in section 3.3 are the same ($r = 0.6$ in the GA).

The simulated arm consists of a forearm, an upper arm (each two units long) and a spherical hand (figure 4.1, (A)). The six target points are spread evenly with uniformly distributed noise $\in [0, \frac{1}{6}\pi]$ on the circumference of a circle with a radius of 1.25. The hand position is always at the middle of the circle, which corresponds to both the shoulder and the elbow angle starting at $\alpha_{e,s} = 60°$ (figure 4.1 (B); in both the two- and the three-dimensional version, the arm also starts with the elbow in the plane).

Both joints are controlled by applying a torque M_i to the joint α_i. In order to test the effect that the number of degrees of freedom (DoFs) has on the task, experiments are run on a planar (i.e., two-dimensional) condition where both the elbow and the shoulder joint have one DoF (α_e and α_s) and a three-dimensional condition, in which the elbow joint has one DoF (α_e) and the shoulder joint has, just like the human shoulder, three DoFs: rotation in the horizontal plane (α_{s1}), lifting/lowering the arm(α_{s2}) and rotation along the arm direction (α_{s3}), figure 4.1 (A)). The arm is constraint by joint stops that follow the human example. Dry friction is applied at all joints. The networks have sensory neurons for the angular position of each DoF and one sensory neuron for the required pointing direction $\phi \in [0, 2 \cdot \pi]$.

There have been some simplifications included into the model that make the agent dynamics very much unlike natural arm movement. In the three-dimensional environment, it is very difficult for evolution to keep the hand close to the plane, something which is automatically afforded by the two-dimensional environment. However, part of the objective of this simulation was to compare a two- and three-dimensional version of the same task. Therefore, the movement in the three-dimensional condition has been constrained such that the hand cannot to deviate from the horizontal plane, meaning the possible hand trajectories are equal between the two conditions, but having more motor redundancy in the three-dimensional version. However, this restriction

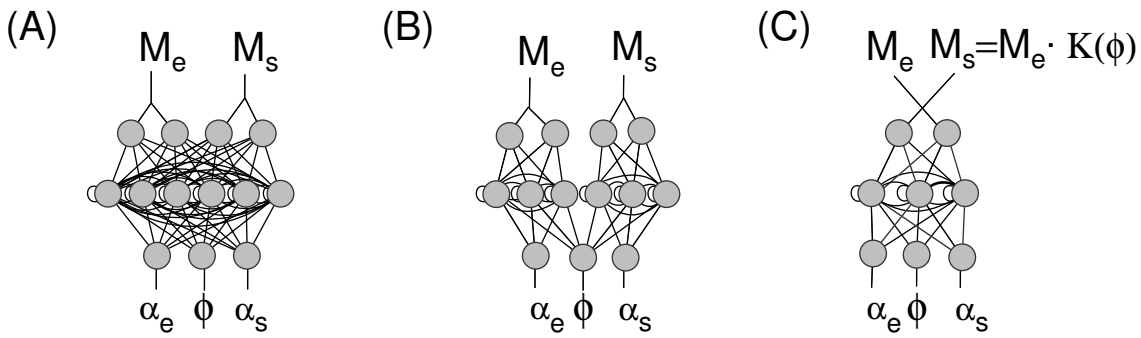

Figure 4.2: Network diagrams for the unconstrained (A), the modularised (B) and the forced synergy (C) condition.

makes the movement more like moving an object across a surface than like natural human reaching movements. For similar reasons, gravity has not been modelled. These constraints reduce biological plausibility of the model and probably make it less suitable for generating quantitative hypotheses for further real experiments. The principal idea, however, i.e., to explore the questions of redundant DoFs in a motor control task, is preserved upon introduction of these additional constraints.

The weights of the CTRNN controllers evolved in the ranges $w_{ij} \in [-7,7]$, the bias $\theta_i \in [-3,3]$ and the time constant $\tau_i \in [0.1,1.77]$[3] with a simulation time step of 0.01, which is the same time step used in the ODE simulation of the environment. Other parameter ranges are $M_G \in [0.1,30]$ and $S_G \in [0.1,20]$. Other than in most simulations, M_G and S_G were evolved individually for each DoF.

Four different neural controllers were evolved and compared for both the two-dimensional and the three-dimensional conditions. In the condition that I call the *unconstrained* condition, a monolithic CTRNN with six hidden nodes per DoF and two output neurons for each torque signal $M_i = M_G(\sigma(a_{Mi1}) - \sigma(a_{Mi2}))$ is evolved (see network architectures in figure 4.2 (A)). In the *modularised* CTRNN has the same number of neurons, but connectivity is decreased, such that two sub-controllers generate the motor signals for each joint individually (see figure 4.2 (B)). They have three hidden neurons each and receive only proprioceptive input only for the joint they control. However, they share the directional task input neuron. Note that in the three-dimensional condition, one of the sub-networks generates three motor signals. Comparing results from the unconstrained monolithic and the modularised condition is interesting with respect to the question of *neural basis of motor synergies*. In principle, coordination between joint movements could well be mediated through the environment and result from the task dynamics. If synergies emerge despite the absence of neural connections between control modules that generate motor signals for each joint, exploiting closed-loop sensorimotor dynamics, such regularities pose a challenge to homuncular explanations of synergies.

In the third and fourth condition investigated, a linear relation is imposed between torques

[3]During write-up, I realised that the presentation of these results (Rohde & Di Paolo, 2005) has a mistake; the minimal τ is given as 0.01, which would be the same as the time step and dynamically unstable. The present value $\min(\tau) = 0.1$ is the actual value that I retrieved from the code of the original experiment.

Table 4.1: Number of parameters evolved.

	unconstrained	modularised	Forced synergy (linear)	Forced synergy (RBFN)
2D	109	75	53	46
3D	161	115	62	83

applied to the elbow joint α_e and the different DoFs in the shoulder α_{si}. I refer to this type of controller as *forced synergy* controller. In these networks (see figure 4.2 (C)), M_e is generated by a CTRNN with three hidden nodes and all joint inputs. The other joint torques M_{sj} are scaled as a linear function $M_{sj} = K_j \cdot M_e$ where K_j is constant within a pointing movement, but varies systematically across trials with the desired pointing direction: $K_j = f(\phi)$.

Two different functional representations are used for the forced linear networks. In the *linear* forced synergy condition $K_j(\phi)$ is a simple linear function for each DoF j

$$K_j(\phi) = k_j^1 \cdot \phi + k_j^2 \tag{4.1}$$

with $k_j^i \in [-4,4]$ set genetically.

The more complex representation of the linear synergy function $K_j(\phi)$ as a RBFN is motivated by the fact that RBFNs are generic representations of continuous functions of the angle, i.e., it does not have a singularity at $\phi = 2\pi$ like equation (4.1). In the RBFN condition, $K_j(\phi)$ is represented by a Radial Basis Function Network (RBFN) with Gaussian RBFs

$$K_j(\phi) = \sum_{i=1}^{4} w_{Ri} \cdot e^{-\frac{\delta^2}{2 \cdot \Delta^2}} \tag{4.2}$$

where $\delta = c_i - \phi, d \in [-\pi, \pi]$ is the difference in direction between the evolved RBF center $c_i \in [-\pi, \pi]$ and the target direction ϕ. The width of the Gaussian RBF $\Delta \in [0.5, 1.5]$ and the RBFN weights $w_{Ri} \in [-4,4]$ are also evolved. The absolute values of the coefficients $|k_i|$ and the absolute values of the RBFN weights $|w_{Ri}|$ are mapped exponentially.

The number of parameters evolved in each condition varies between 46 and 161 (see table 4.1).

Trials are run for $T \in [2000, 3000]$ time steps. The fitness $F_j(i)$ of an individual i on a target spot j is given by

$$F_j(i) = 1 - \frac{d_j(T,i)^2}{d_j(0,i)^2} \tag{4.3}$$

where $d_j(t,i)$ is the distance of the hand from the target spot j at time t for individual i.

Networks for all conditions are evolved with either incremental evolution (i.e., they were evolved on just a sub-set of target spots, starting with two target spots, and the next clockwise target spot is added to the evaluation once the average performance of the population exceeds $\bar{F} = 0.4$) or on all six target spots right from the start. The evaluation of a network i on n target spots is calculated using the exponentially weighted fitness average defined in section 3.3, equation (3.4).

4.3 Results

The results presented in this section focus on several of the aspects investigated. Evolvability is a variable that plays an important role throughout this section and is, mostly, indicated as the number

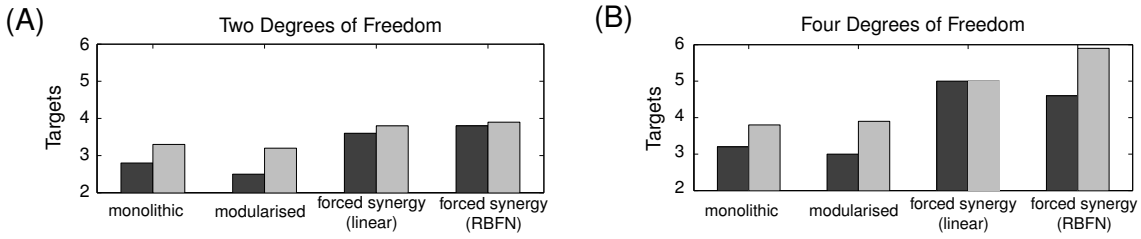

Figure 4.3: Average number of starting positions reached in incremental evolution after 100 (dark) and 500 (light) generations across ten evolutionary runs.

of target spots the network was evolved on in incremental evolution, as this variable corresponds to grades in performance. Subsection 4.3.1 compares the two- and three-dimensional version of the simulation across neural controllers focussing on the role of spatial redundancy. Subsection 4.3.2 compares the results from the different kinds of network controllers, focusing on the role that forcing a motor synergy plays for evolvability and performance. The last subsection 4.3.3 takes a closer look at the linear synergy functions $K_j(\phi)$ evolved to solve the task.

4.3.1 Number of Degrees of Freedom

The problem of *motor redundancy* identified in the introduction already applies in the two-dimensional version of the task, because there are infinitely many trajectories to move the hand from position P_A to position P_B. However, for any position P, in this set-up, there is (due to joint stops) just one possible pair of joint angles (α_e, α_s) to realise it. In the three-dimensional set-up, due to the three DoFs in the shoulder joint, there are infinitely many shoulder positions $\alpha_{1,2,3}$ associated with a position P, even if the elbow angle α_e is not redundant. The space of motor signals to arrive at a configuration is even more redundant, due to the fact that the network generates torques, rather than angular velocities or joint positions, so different interfering forces (passive dynamics, interaction of torques applied to different joints through the body and the environment) work on each joint and affect the arm trajectory.

Averaged across 10 evolutionary runs, the motor redundancy afforded by the three-dimensional set-up provided a clear advantage in evolvability (see figure 4.3) in all network architectures: in the incremental evolution condition, the number of target spots reached is much higher.

Exploring the space of strategies evolved in case studies, I found a much greater variety of solutions in the three-dimensional version than in the two-dimensional version, where the only variation in strategies to reach a certain target spot is to temporally vary the torques applied to both joints in order to bring the two planar joints in the appropriate end positions. The motor redundancy afforded by including two additional DoFs in the shoulder joints, on the other hand, allows for a greater variety of strategies that exploit the additional DoFs and environmentally mediated forces. Among the monolithic and the modularised CTRNN controllers, a common strategy is to turn the arm along its length to one of the joint stops, leaving the hand in the centre of the plane, before moving to the target spot. It seems that the positions thus reached are more suitable for evolutionary search and directional reaching than the original starting position.

To gain further insight into the mechanisms and robustness of the evolved solutions, I in-

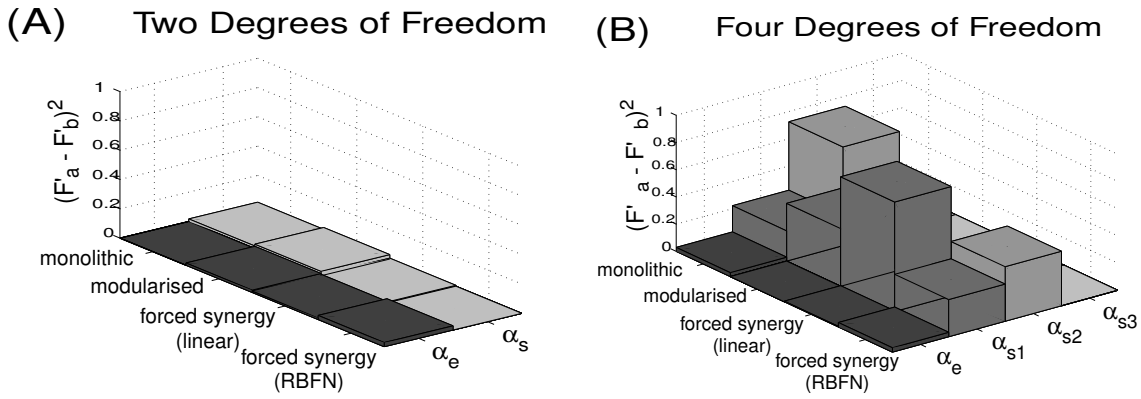

Figure 4.4: Squared difference in normalised performance as individual joints α_i are free to move but not driven (F'_a) or blocked (F'_b) in example two-dimensional (A) and three-dimensional (B) agents evolved.

vestigated how the controllers reacted if individual DoFs were disabled. To investigate the role of passive dynamics, I compared conditions: in condition $F'_a(i)$ individual DoFs were 'anaesthetised', i.e., passive dynamics were possible, but no motor torques were applied. In condition $F'_b(i)$, individual DoFs were blocked, i.e., the joint angles were fixed at their initial position. Figure 4.4 shows the squared difference in performance $(F'_a(i) - F'_b(i))^2$ between these two conditions per DoF affected and network type. In the two-dimensional condition, enabling passive dynamics to work on the anaesthetised DoFs hardly make a difference in performance (figure 4.4 (A)), whilst in the three-dimensional condition (figure 4.4 (B)), it has a noticeable impact on performance of all networks. The controllers evolved in the three-dimensional set-up, therefore, appear to make use of the motor redundancy and increased possibilities for passive dynamics in order to increase stability of the solution.

These findings from the simple simulation models show how in a sensorimotor task the inclusion of additional DoFs can increase evolvability. In the pursuit of minimalism, it is tempting to endow an agent with the minimally required sensorimotor system for a task but such an idealisation can introduce a bias into the sensorimotor dynamics, delimit the strategies evolved and hamper evolution of high performing solutions, despite the reduction of the search space (cf. table 4.1).

4.3.2 Forcing Linear Synergy

Comparing the evolution of an unconstrained monolithic or modularised CTRNN controller with the networks that were evolved to act in linear synergy, we find that the agents in whose control architecture linear synergies are built-in reach much higher levels of performance on average, both in the two-dimensional and in the three-dimensional condition. Figure 4.3 depicts the number of target spots that each network type evolved to solve in the incremental evolution condition. The RBFN synergy networks advance to the next goal twice as many times as the networks that are not imposed this constraint. With twice as many generations, the CTRNN controllers without forced linear synergy come close but never reach the level of performance of the networks forced to act in linear synergy.

The only agents that evolve to solve the entire problem space are the RBFN synergy agents in the three-dimensional set-up; in all other conditions, evolution stagnates in a sub-optimal level of performance on a limited number of target spots, such that the population average does not exceed 0.4 to enter the next stage of incremental evolution. In the three-dimensional forced synergy condition, average performance of best individuals after 1000 generations is 0.65. Non-incremental evolution led to qualitatively similar results, i.e., quicker and more successful evolution of forced synergy networks, even if, quantitatively, the overall fitness evolved was much lower in a non-incremental approach.

It may be argued that choosing RBRNs to represent variation in directional scaling is a biased choice that builds in part of the solution already. I think this is arguable in the case of the RBFN, but it is certainly not the case for a simple linear function, which has a singularity at $\phi = 2\pi$. The two-dimensional scenario is already very restricted. Forcing linear synergies to relate so crudely to the required pointing angle makes it impossible to generate a controller that masters the task. Despite these principal limitations, the solutions for all set-ups in which networks were forced to act in linear synergy evolved to much higher levels of performance.

To rule out the possibility that the simple CTRNN controllers (monolithic or modularised) could not cope with the presentation of the input direction as a scalar neural input, a more 'CTRNN friendly' set-up was tested, where controllers were provided with six different input neurons for the different target spots and no noise applied to ϕ. Still, neither in the two-dimensional nor in the three-dimensional condition did the agents advance beyond the presentation of three target spots within 1000 generations. There seems to be something about functionally dividing the task into the generation of a torque signal and determining separately how this torque signal is divided between the different DoFs that is particularly suitable for artificial evolution to efficiently evolve good solutions for the given task.

4.3.3 Evolved Synergies

The last analysis was the kind of linear synergies $K_i(\phi)$ which evolved. However, no general pattern could be observed. Figure 4.5 depicts example RBFN synergies evolved for the three-dimensional condition: the overlap and difference in centre of the peaks in these RBFs, however, explains the diversity of behavioural strategies for different ranges of ϕ observed in the RBRN agents: For different targets, different DoFs are predominant in the realisation of the task.

Imposing linear synergy increases evolvability of solutions. A possible explanation for this increase in evolvability is that such solutions are particularly suitable for realising the motor task evolved. If this was the case, an increase of linearity in torque relation could be expected as a result of evolutionary advance in the (monolithic and modularised) CTRNN controllers. Figure 4.6 (A) shows the sum of squared error from linear synergy in these types of networks changes across evolution in the best individuals evolved for both the two- and the three-dimensional condition (average across five evolutionary runs). In the two-dimensional condition, there is a tendency to reduce this error, i.e., to get closer to linear synergy, as performance increases. In the agents evolved for the three-dimensional networks, in contrast, linear synergy and performance appear to be completely unrelated.

The modularised CTRNN controllers are on average much less prone to exhibit linear synergy

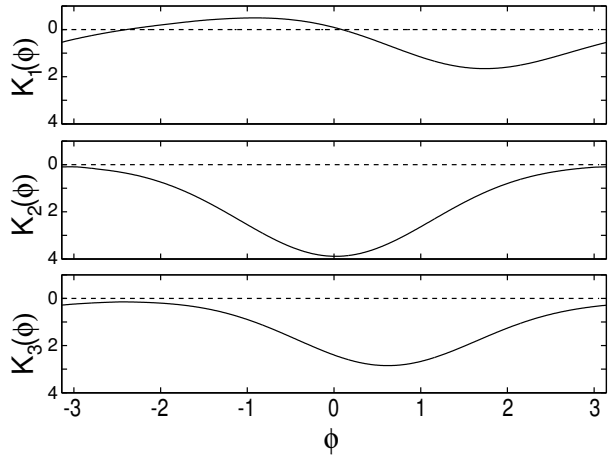

Figure 4.5: An example evolved RBFN for a forced synergy network for the three-dimensional condition.

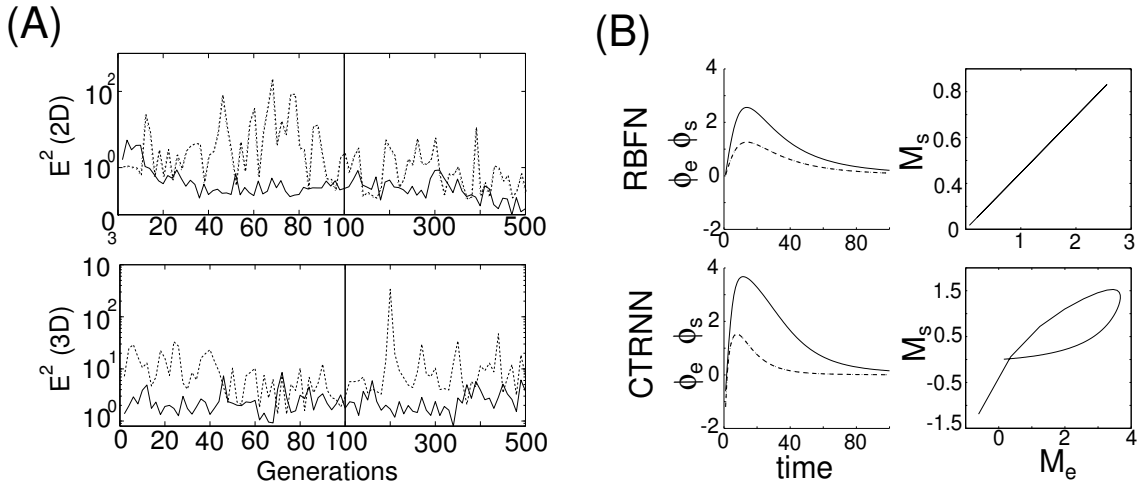

Figure 4.6: (A) Sum of squared deviation from linear synergy across generations in the two-dimensional (top) and three-dimensional (bottom) networks. Solid: unconstrained, dashed: modularised; average across five evolutions. (Note non-linear scales) (B): A two-dimensional monolithic CTRNN controller (bottom) applies a similar strategy as a RBFN forced synergy controller.

(note logarithmic scale), even there is a lot of variance in this variable. The reason to investigate this network architecture and compare it to the monolithic CTRNN controllers was that, if linear synergy was a generically good strategy in the task, this relation between the joint torques could have been implemented even without a neural structure controlling it, instead exploiting the environmental dynamics to achieve coordination. Given the exploitation of the environmental for joint control using passive dynamics described in the previous section 4.3.2, it is clear that the simulation used does, in principle support exploitation of environmental dynamics. However, linear synergies without a neural basis did not evolve. Also, being more disposed through neural connections to coordinate joint torques does not appear to provide the monolithic CTRNN controllers with an evolvability advantage (cf. figure 4.3 (A) and (B)). All these findings suggest that the magnitude of deviation from linear synergy is not an essential characteristic of a successful solution.

Figure 4.6 (B) shows how a monolithic CTRNN controller in the two-dimensional scenario applies a very similar strategy as a controller that is forced to act in linear synergy. The CTRNN controller emits motor signals to the two joints with a slight delay, as also represented by the loop in the M_e/M_s map. Such temporal displacement disrupts linear synergy as defined in this paper, but, on a functional level, this does not mean a disadvantage.

4.4 Discussion

The results of the series of ER simulations conducted, even though the model abstracts strongly from the human original, produces some interesting proofs of concept and hypotheses for further empirical experimentation. Firstly, linear synergies could not be found to be the outcome of an unconstrained evolutionary search process. Also disconnecting controllers for different joints did not provide a disadvantage in evolvability compared to monolithic networks controlling both joints. This suggests that the mere possibility of implementing systematic relationships between effectors in a network does not provide a selective advantage.

On the other hand, imposing the constraint of linear synergy strongly improves evolvability of viable solutions, even if the function $K_j(\phi)$ that specifies the relation between the joint torques is a simple linear function (equation (4.1)), but even more so if this relationship is represented as a RBFN (equation (4.2)) that allows to define more complex and continuous functions of angles. The division of control into scaling function and generation of motor signal is suitable for evolutionary search in the given task. It is, however, unclear what exactly this benefit consists in. I analysed the ruggedness of the fitness landscape around successfully evolved individuals by applying random mutations of increasing magnitude r to compare the decay in performance. This test can indicate the slope and ruggedness of the local fitness environment. No noticeable discrepancies between the different conditions could be shown, and the decay profile between controllers within the same condition varied considerably at a comparable level of performance.

Arguably the most interesting result from this model is that both a complication of the parameter space (i.e., adding more DoFs) and a simplification of the parameter space (i.e., forcing linear synergy) have provided independent evolutionary advantages. Thinking of the search space in numbers of parameters evolved (table 4.1), it turns out that both the best configuration (three dimensions, RBFN synergy) and the worst configuration (two dimensions, monolithic or modu-

larised CTRNN) are in intermediary range of evolved parameters. Thus, improving evolvability is not a matter of scaling up or scaling down the search space, but of *reshaping the fitness landscape*. As tasks and robotic platforms become more complex, Evolutionary Robotics must produce appropriate reshaping techniques to scaffold the search process and thereby solve the 'bootstrap problem' (Nolfi & Floreano, 2000, p. 13) and biology may be a suitable source of inspiration in searching such appropriate constraints.

The fact that both the monolithic and the modularised CTRNN controllers failed to evolve linear synergies suggests that this organisation of movements is not as such beneficial in the given task. The dramatic increase in evolvability that imposing linear synergies onto the movement space means proposes an explanation that is more in line with Zaal et al.'s (1999), i.e., that constraining the space of solutions by imposing linear synergy is a beneficial pruning of the space of behavioural possibilities for a developmental process (artificial evolution or motor development) to learn efficiently. In order to further investigate this hypothesis, it would be interesting to study the phylogeny of linear synergy, or, as an extension to the experiments presented here, to evolve the constraints for ontogenetic development, hypothesising that linear synergies would result from unbiased evolution in this meta-task.

Another interesting finding is that in the three-dimensional simulation, passive dynamics and redundant DoFs could be shown to be exploited, whereas in the two-dimensional version, the solutions evolved appeared to be less sensitive to environmental forces. The restriction of movement to the plane constrains behavioural possibilities much more drastically than imposing linear synergies between joint torques.

It has to be stressed that the results about the beneficial role of linear synergies do not automatically generalise to all kinds of tasks. In contrary, it is quite obvious that, for instance, a two-wheeled robot doing obstacle avoidance (a simulation which is not redundant in DoFs) will rely on an ongoing change in the relation between the effectors. There is, however, a possible analogy to be drawn to physiological data again: as mentioned in the background section 4.1, evidence from studies on human physiology suggests that linear synergy can be broken. Probably such a deviation from this unlearned principle of motor organisation is acquired if such variability in the relation between actuators serves the task.

As concerns the scientific value of ER simulation models for the study of human behaviour and cognition, the present model has generated some proofs of concept and hypotheses that help in theory building and the design of further experiments. Findings on systematicities between effectors, as they are ubiquitous in humans and animals, have been explored with an ER simulation model to investigate their function in an unbiased way. Even though the exact functional role of motor synergies remains a mystery, the concept of motor synergies, even if it derives from a homuncular view on motor control, appears to be a useful concept that can be integrated into a more enactive story of motor control.

A concern of mine had been that the results, even though they may be in principle relevant for experimental scientists in the field of human motor control would not be perceived or acknowledged by the research community. In that case, this model would have only been useful to the extent that it advances the field of artificial life and generates heuristics to evolve artificial agents efficiently, which was not the primary research goal. Approaching the key researchers of both

experimental studies that inspired my work (Gottlieb and Zaal), I was very happy to find acknowledgement of the significance of my work by both. The Gottlieb group refer to the conference paper in which these results were published in a paper about the effect of movement direction in joint torque co-variation (Shemmell, Hasan, Gottlieb, & Corcos, 2007) that follows up to their original study. They write

> "Results obtained from a study in which a simulated robotic arm was evolved to reach a number of target locations appear to support this conclusion (Rohde and Di Paolo 2005). The results of the simulation demonstrated that linear synergy was not a control solution converged upon by an unconstrained (i.e., not forced to produce temporally coupled torques at both moving joints) neural network in order to reach the designated targets. The same simulation however, showed that the imposition of linear synergy as a kinetic constraint significantly improved the ability of the neural network to evolve and reach the designated targets" (Shemmell et al., 2007, p. 157).

Zaal acknowledged the work in personal (email) communication, recognising its contribution as well as the potential of the technique for further research: remarking that the group had been criticised for the fact that the gravitational component of the torque had been left out in the calculation, he proposed to explore the role of this component in further simulation models.

This proposal by Zaal is in line with my own ideas of how the presented research could have been extended to gradually increase biological plausibility and, thereby, comparability with humans. Apart from introducing gravity into the model, a good modification of the model would be to allow for the hand to deviate from the plane in the three-dimensional version of the task, which would immensely increase similarity of the simulated task to the original task. Extending the model this way would require the networks to additionally solve the non-trivial task of equilibrating forces involved to counter gravity and deviation from the plane.

I chose, however, not to extend this promising line of research from the first year of my doctorate research. Even though I do think that the research question addressed is relevant for Cognitive Science, problems of motor control were very remote from high-level cognition and human experience. As outlined in chapter 2 section 2.4, embodied and dynamical approaches are sometimes criticised to be confined to such low level behaviour - motor synergies are not a computationalist stronghold, a presumed 'representation hungry' problem. The following modelling and experimental chapters present results from models that address questions that are much more at the centre of Cognitive Science and higher levels of cognitive organisation.

Chapter 5

ER Can Illustrate and Verify Conceptual Arguments: An Exploration of Value System Architectures in Simulation

---

> "Our desires and our projects are conditioned by certain needs of our nerves which are hard to define in words"[1]
> Thomas Mann, *Buddenbrooks* (1985, recent edition; original published in 1901)

The previous chapter presented an ER simulation model applied to a problem of motor control, whose results immediately suggest further empirical experiments of the kind that had been modelled. This very tangible way of using ER simulation models in dialogue with empirical sciences directly relates to its capacity to prove dynamical principles and to take logics and mathematics beyond our cognitive grasp of complex interaction dynamics, as explained in chapter 3 section 3.3. In this chapter, I present a simulation model whose results are of a more abstract and conceptual nature. It investigates the conceptual soundness of arguments proposing a certain type of neural architecture (value system architectures) as explaining life-time adaptation. The model presented in this chapter is followed by the first applications of ER modelling to PS experiments (perceptual crossing in one dimension in chapter 6 and in two-dimensions in chapter 7), which aims to combine both merits of ER modelling, as argued in chapter 3 section 3.6. These studies lead to the second part of this dissertation (chapters 8-12) which presents the results from the interdisciplinary study of adaptation to sensory delays and perceived simultaneity, for which I conducted both the experimental and the modelling work, before the conclusion from the entire body of work presented here is discussed in chapter 13.

The results from the model presented in this chapter have been partially published in (Rohde & Di Paolo, 2006c; Di Paolo et al., 2008a). I model a very wide-spread cognitive architecture in which life-time adaptivity is thought to be driven by a reinforcement signal, generated by a structurally isolated and functionally dedicated module. I call these architectures 'value system architectures', adopting the terminology of Edelman et al. (e.g., Sporns & Edelman, 1993), but the idea is much more common than just the research presented under this label. My simulation model illustrates some of the implicit premises that underlie this kind of architecture by demonstrating

[1] My translation: "Unsere Wünsche und Unternehmungen gehen aus gewissen Bedürfnissen unserer Nerven hervor, die mit Worten schwer zu bestimmen sind" (Mann, 1985, recent edition; original published in 1901).

how, without building in additional constraints, the adaptive capacity of such circuits can break down under neural (or physiological, environmental) plasticity.

In the background section 5.1, I introduce not only the question addressed my understanding of value system architectures (as described in (Rohde & Di Paolo, 2006c)): I also develop some ideas on sense-making in the enactive approach that were briefly mentioned in chapter 2 and which have been published as part of (Di Paolo et al., 2008a). The model and its results are presented in sections 5.2 and 5.3. The discussion section 5.4 evaluates the results with respect to the question previously framed and with respect to the methodological theme of the current dissertation. This section also identifies possibilities for future simulation studies to complement the results presented here.

5.1 Background: Enactive and Reductionist Approaches to Value

In this section, I elaborate on the idea of sense-making as a core concept in the enactive approach as it is sketched in chapter 2 section 2.3 and developed in more detail in (Di Paolo et al., 2008a). I then outline how this idea of sense-making is in tension with reductionist views and modularisation of function, before I map the outlined debate on the issue of value system architectures and present the question addressed with the simulation model.

5.1.1 Sense Making, Value Generation, Meaning Construction

Weber and Varela (2002) have been the first to explicitly identify intrinsic teleology, natural purposes and the possibility to imbue interactions with the environment with meaning as fundamental properties of living creatures. They combine ideas from Kant's 'Critique of Judgement' and Jonas' biophilosophy (1966) in order to argue that autopoietic organisation does not only imply basic autonomy and identity generation (as it had been argued before (e.g., Maturana & Varela, 1980)) but also implies genuine purposefulness of existence and interactions with the environment. Thereby, a profane material process obeying the laws of physics, like two celestial bodies colliding or water streaming down a mountain, becomes meaningful and can be positive, negative or ambivalent to the organism, depending on its impact on autopoietic organisation. I generally commit to this idea and I want to stress that it relates to my work on autonomy and generative mechanisms presented in (Rohde & Stewart, 2008, cf. chapter 4 section 3.1). It also relates to Di Paolo's interpretation of Weber and Varela's ideas in the context of autonomous robotics (Di Paolo, 2003) and his extension of their ideas (2005), in which he argues that adaptive autopoiesis is the minimal condition for sense-making, whereas 'passive' autopoiesis only generates purposes.

We propose to define value as "*the extent to which a situation affects the viability of a self-sustaining and precarious process that generates an identity*" (Di Paolo et al., 2008a). Autopoiesis, i.e., the continued self-construction of a metabolising network of processes sustaining itself in a far-from-equilibrium situation (which characterises life) is the most prominent example of such a process. The simplest form of autopoiesis, a single celled organism (such as a bacterium) serves best in order to argue how encounters with the environment can be called good or bad depending on their consequence for continued autopoiesis.

This basic 'metabolic value' is, however, not the only process that is self-sustaining, precarious and generates an identity. More complex forms of organisation give way for multiple levels of such

identity generation and, consequently, to different values which may not relate to metabolism or even generate a conflict in opposing the basic metabolic needs of the organism. In the remainder of this section, I want to present some preliminary ideas on this important and largely undeveloped research question.

Varela explored the idea of the organism as a 'meshwork of selfless selves' (1991) and 'patterns of life' (1997) in several places, identifying how in phylogenetically more developed organisms new levels of autonomous dynamics can emerge on top and alongside autopoiesis. His own scientific work focussed on three such levels of autonomous dynamics: autopoiesis (cellular identity), the immune system (multicellular identity) and the nervous system (neuro-cognitive identity). Varela identifies other levels of possibly identity generating processes, reaching from pre-cellular identity (self-replicating molecules) to socio-linguistic identity and superorganismic identity.

In (Di Paolo et al., 2008a), we have developed what we call the 'scale of mediacy' (figure 5.1). It lists a number of transitions in value-generating mechanisms that we find particularly important, drawing on the just summarised ideas by Varela (1997, 1991) and a similar analysis of levels of organismic qualitative complexity developed by Jonas (1966). Similarly, Stewart (2008) identifies further important transitions in the phylogeny of cognition.

The first three stages of this scale are not usually identified as distinct. However, the above mentioned work by Di Paolo (2005) distinguishes mere autopoiesis from adaptive autopoiesis, in which the recognition of environmental tendencies and according reactions form the basis for generating value and meaning that goes beyond just life and death. The distinction of the third level, i.e., of interactive regulation is based on Moreno and Etxeberria's (2005) observation that regulation of internal state only cannot be justly called agency. In order to call a living organism an agent, they argue, it has to also adaptively act on the environment. "An example of a just-adaptive organism is the sulphur bacterium that survives anaerobically in marine sediments whereas bacteria swimming up a sugar gradient would, by virtue of their motion, qualify for minimal agency" (Di Paolo et al., 2008a). The further stages that are included in the scale (figure 5.1) have been adopted from Jonas' (1966) work. He identifies the fast motility of animals as the basis of emotions because they can assign meaning to something at a distance and thus fear or desire the remote. The last two stages are reserved to humans, who, through their general image-making capacity, and particularly their self-image-making capacity, gain the ability to regard situations objectively and define themselves as subjects.

This listing follows the 'gradient of mediacy' in that the meaning for the precarious process and its sign to the organism become increasingly mediated and physically detached. The consequence of increased mediacy is the liberation of ways to generate values: "for instance, only a sense-making organism is capable of deception by virtue of the mediacy of urge and satisfaction. A bacterium that swims up the 'saccharine' gradient, as it would in a sugar gradient, can be properly said to have assigned significance to a sign that is not immediately related to its metabolism, even though it is still bound to generate meanings solely based on the consequences for its metabolism" (Di Paolo et al., 2008a). The higher the degree of mediacy, the more complex it is for the observer to interpret a sign with respect to the process(es) of identity generation from which its value emerges. Therefore, the meaning that animals, and particular human animals assign to most of

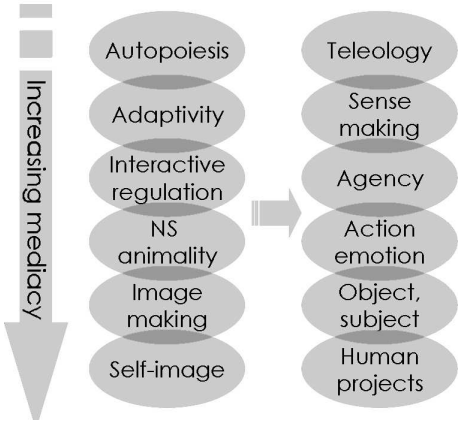

Figure 5.1: Life-cognition continuity and the scale of increasing mediacy.

their interactions with the world are by no means easy to explain and can only very indirectly, if at all, be related to metabolic self-production. This insight resonates, to some extent, with Barandiaran's (2008) concept of 'Mental Life' as an entirely different level of autonomy than metabolism.

The ideas developed in this section are not meant to provide a full-blown theory of the origin of values; they should, however, give an impression of the kind of question, method and answer the enactive approach adopts towards problems of meaning, value, purpose, intentionality, autonomy, emotion, etc.: The inherent meaning in a process is investigated in its situated and embodied context, with reference to biological, evolutionary and dynamical evidence and principles.

This enactive study of value involves the study of generative mechanisms, as it had been argued for the study of autonomy (chapter 3 section 3.1). Therefore, the study of the major transitions in evolution (such as the evolution of the eukaryote cell, of sex or of multicellularity) that is described by Maynard Smith and Szathmáry (1995) is important in complementing the study of the scale of mediacy. New forms of organismic organisations can enable new and more complex kinds of value-generating processes, even though I do not presume a one-to-one mapping between transitions in structure and transitions in value-generating processes. I hypothesise that the latter proceeds much more gradually.

All these are exciting avenues for future research and not settled issues. Another open research question is whether there can be value-generation according to my definition without the need of an underlying metabolism as basic level of autonomy and how different levels of values are integrated or generate conflicts.

5.1.2 Reductionist Approaches to Values and Value System Architectures

In representationalist approaches the symbol is separated from its meaning - the *signifiant* from the *signifié* - processes are not *inherently* meaningful in the way described above, but are syntactic and become interpreted, as explained in chapter 2. The question of the origin of values thus has to be approached in a fundamentally different way, looking for a process or entity *external* to the syntactic 'cognitive' process itself that provides meaning for the computational tokens. Many reductionist approaches refer to natural selection and survival of the fittest in Darwinian evolution as inherently purposeful process that ensures that information processing is set up in a way that

promotes genetic proliferation (e.g., Millikan, 1984): behaviour is meaningful only in so far as we can explain how it helped our ancestors to survive and reproduce in the African savannah.

This extreme reductionist perspective just sketched can be seen as one pole of a spectrum, in which a purpose precedes the living organism, a concept which I call *a priori semantics*. This pole is in strong opposition to the enactive approach described above, in which evolution is an essential factor *shaping* the levels of mechanical processes that generate meaning but does not provide meaning itself. There are intermediary positions between these two poles that try to follow a third route, assuming that some, but not all meaning is determined evolutionarily.

A strong group in the intermediary group are the proponents of what I call 'value system architectures' adopting Edelman et al.'s (e.g., Sporns & Edelman, 1993) terminology from the Theory of Neuronal Group Selection (TNGS). The kind of architecture discussed is, however, much more widely used than this label: what I call value system architectures are all those models that assume that parts of the cognitive/neural architecture are functionally and structurally isolated from the behaviour generating and plastic parts of the architecture and that those encapsulated modules represent and provide (evolutionarily determined) meaning.

In the concrete proposal of TNGS, these value systems are presumed to generate a bipolar performance signals that evaluates sensorimotor behaviour. Sporns and Edelmann define value systems as neural modules that are "already specified during embryogenesis as the result of evolutionary selection upon the phenotype" (Sporns & Edelman, 1993, p. 968). Such internally generated reinforcement signals direct life-time adaptation ('value-guided learning'): a value system for reaching, for instance, would become active if the hand comes close to the target. A functional and structural division between behaviour-generating mechanisms and mechanisms of value-based adaptation is at the core of value system architectures.

This kind of architecture is very popular with skeptics of the traditional paradigm who argue for more embodiment and situatedness. For instance, Sporns and Edelman see this kind of an architecture as a solution towards problems of anatomical and biomechanical changes that are described as "challenging to traditional computational approaches" (Sporns & Edelman, 1993, p. 960). Pfeifer and Scheier, two pioneers of the situated and embodied approach in AI, argue that "if the agent is to be autonomous and situated, it has to have a means of 'judging' what is good for it and what is not. Such a means is provided by an agent's value system" (Pfeifer & Scheier, 1999, p. 315) and present Verschure et al.'s (Verschure, Wray, Sporns, Tononi, & Edelman, 1995) implementation of a TNGS architecture as the way forward in autonomous robotics. Similarly, Varela's co-author Weber, who is cited above for first linking autopoiesis and organismic autonomy to genuine purposes, praises TNGS in his biosemiotic theory (Weber, 2003).

I believe that the idea of value system architectures is conceptually problematic in assuming separation of behaviour and value, in reducing function to local mechanism and in the teleonomical assertion that evolution builds in values. This argument is largely in line with and partially based on a similar but more instrumental or pragmatic criticism of value system architectures by Rutkowska (1997) who argues that "[increased] flexibility requires some more general purpose style of value" (Rutkowska, 1997, p. 292) than a value module could provide. She believes that value system architectures cannot explain adaptivity as a general phenomenon, even if value-guided learning circuits may work in specific cases. As we argue

> "She laments their vulnerability and their restrictive semantics consequent to the built-in evaluation criteria. A similar limitation is pointed out by Pfeifer and Scheier, who describe a 'trade-off between specificity and generality of value systems' (Pfeifer & Scheier, 1999, p. 473): A very specific value system will not lead to a high degree of flexibility in behaviour, while a very general value system will not constrain the behavioural possibilities of the agent sufficiently" (Rohde & Di Paolo, 2006c).

Rutkowska goes as far as posing the question as whether a value system is a "vestigial ghost in the machine" (Rutkowska, 1997, p. 292).

Drawing a box and labelling it 'value system' seems an approach very different from the enactive exploration of dynamical processes of identity generation and the significance they bring about. The kind of reasoning associated with value system architectures bears traces of homuncularity and boxology in assuming that what is good can be specified and represented as a function *pre factum*. As such, value system architectures suffer, in a miniature version, from those problems identified to result from the computationalist paradigm in chapter 2, section 2.1: rigidity, semantic limitations, incapacity to deal with open-ended real-time change, etc.

Is could be that the above listed researchers, despite seeming so close in to the enactive approach, are really just closet computationalists. I do not believe that this is so. Whilst I do not believe in paradigmatic middle grounds (as argued in chapter 2 section 2.2), I think that there are certainly models and results that can be interpreted either way. Decades of exercising a computationalist methodology persist in the language used to formulate questions and this makes it very difficult to fully let go of the baggage of implicit premises. It requires a constant attention to such issues to avoid postulating vestigial ghosts in the machine. Nobody disputes that norms exist across individuals of a species that result from natural selection. But there is a thin line between arguing that these norms are built in as parts of the mechanism, which is reductionist, and investigating the mechanism that give rise to such norms that manifest in the relational and behavioural domain, which is not reductionist. This is a very important but very subtle difference, which many researchers may not be constantly aware of. I believe that the cited promoters of value system architectures are likely to be victims of light paradigmatic confusion. This means that the problems are not so much rooted in the circuits proposed but in calling parts of it a 'value system' and asserting that their meaning is built in by evolution.

It is important, though, to realise that this kind of language is not just a shorthand and actually in line with the enactive approach. It sneaks in further computationalist premises through the backdoor. Edelman's statement that "[TNGS] relies only minimally upon codes" (Edelman, 1987, p. 45) and that "general information about the kinds of stimuli that will be significant to the system is built in" (Edelman, 1989, p. 58) are literally computationalist and reductionist. Similarly, Weber (2003) explicitly endorses the translation of meaningful variables into neural constraints when he writes that "only some vital behavioural emphases are built into the neuronal architecture from birth. They work as basal orientation values. [...] only very few fixed values exist (like, e.g., 'food is good', 'light is good' etc.)" (Weber, 2003, p. 60).[2]

The simulation model described in this chapter illustrates the consequences of such a reduc-

[2]My translation: "Nur wenige vitale Verhaltensschwerpunkte sind bereits bei der Geburt in die neuronale Architektur eingebaut. Sie wirken als basale Orientierungswerte. [...] nur sehr wenige festehende 'Werte' existieren (wie etwa 'Nahrung ist gut', 'Licht ist gut' etc.)" (Weber, 2003, p. 60).

tionist perspective if taken seriously. It thus aims to clarify paradigmatic debate and to resolve paradigmatic confusion.

5.1.3 A Caricature of Value System Architectures in Simulation

The simulation model presented in this chapter illustrates the conceptual argument just made by demonstrating how in value system architectures, the proposed functional separation and localisation can lead to break-down of the proposed adaptive principles in the presence of general neural and behavioural plasticity, if further implicitly held modelling assumptions are not built in. Taking the idea seriously that a local pre-defined structure generates meaning for an otherwise merely syntactic and value-agnostic architecture, it results that there is no way to make sure that the value system keeps working properly, that its input and output channels do not get re-interpreted in a variable sensorimotor context. A value signal that is actually symbolic in that it is arbitrary with respect to the meaning it bears could mean anything and the structures that obey it in performing adaptation has no way of telling what is wrong. Gradual change in meaning through gradual change in sensorimotor context is what I call *semantic drift*, a concept that resulted from the simulation presented below.

ER simulation models are particularly suited to investigate the relation between function and mechanism because this relation is not pre-specified but results from automated search (cf. chapter 3 section 3.3). An example that relates to the problem addressed here is Yamauchi and Beer's (1994) evolution of learning in a fixed weight CTRNN and follow-up work (e.g., Tuci, Quinn, & Harvey, 2002; Izquierdo-Torres & Harvey, 2007). In these simulations, associative learning behaviour is evolved in fixed weight controllers. The motivation of these studies was a skepticism about presuming a one-to-one mapping between functional properties and structural properties, i.e., presuming that synaptic plasticity implements learning whereas neural activation dynamics implements behaviour. The simulation results provide an existence proof that such a functional and structural separation is not *a priori* necessary.

The thrust behind the idea of pre-coded values is based on a similar presumption, i.e., on the idea that there is a pre-specified and behaviour independent isomorphism between the function represented in the value-module and what is genuinely good or bad for the organism, and that value-guided learning modulates the structurally and functionally separate sensorimotor systems top-down. In the tradition of the experiments on fixed-weight learning, the encapsulation of meaningful judgment is critically investigated in the present simulation.

As a neuroscientific theory, TNGS is backed with empirical evidence. Crucially, a correspondence between neural activity of cell assemblies in the brain stem and the limbic system that modulate synaptic changes in the cortex and salient events in the environment has been observed (Edelman, 2003). This neural system is postulated to implement a value system for certain circuits of value-guided learning.

The first simulation model presented (section 5.3.1) evolves agents to perform a simple behaviour (phototaxis) and to evolve an activity signal bearing this characteristic, i.e., to correlate neural activity with behavioural success (fitness), without assigning this value signal any functional role. This is a critical conceptual investigation of what we really can infer about functional localisation from correlated neural activity, without taking on-board further assumptions. The re-

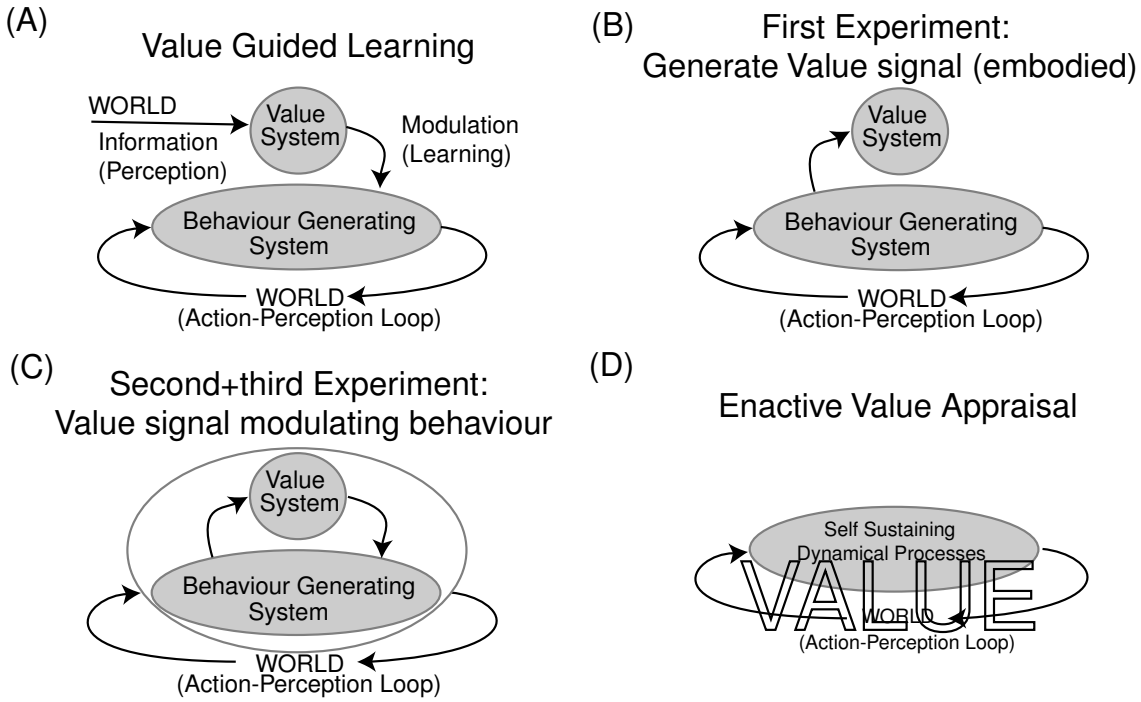

Figure 5.2: An illustration of values (A) in value system architectures, (B) and (C) in the presented simulation models and (D) in the enactive view. (A) Value system function is functionally and structurally separate from the behavioural dynamics and modulates sensorimotor behaviour top-down. (B) In the first experiment, a value signal is evolved to be generated bottom-up in an embodied agent. (C) The second and third experiment incorporate the value system into the brain-body-environment interaction; the second experiment implements the functional reduction proposed by TNGS, the third experiment does not pre-determine the function/structure relation. (D) In a properly enactive view, values are not localised in a neural module but emerge from a meaningful self-sustaining process of identity generation.

sults from this study illustrate how a neural structure with correlated activity can rely heavily on the sensorimotor dynamics of behaviour, rather than to represent external matters of affairs, even if there is no physical connection between this structure and the motor sub-systems.

TNGS proposes ontogenetic Darwinian-style evolution as principle of neural organisation (Edelman, 1989, p. 242). The output of a value system is used to strengthen the synaptic connections participating in the constitution of behaviour that is rated 'good' by the value system, a process akin to natural selection. In the second part of the simulation study (section 5.3.2), the value system evolved in the first part of the simulation study is used to implement the idea of this neural Darwinism. In this particular simulation, artificial evolution is seen as a metaphor of ontogenetic neural Darwinism, not of a phylogenetic process. This model investigates the consequences of the reciprocity of causal links between value system and behaviour generating sub-systems, if sensorimotor dynamics influence value system function. The results show that the behaviour thus evolved does not get better but instead quickly gets worse as a consequence of *semantic drift*.

I want to stress that this way of using a GA and ER simulation as an analogy for neural Dar-

winism is inspired by the neural Darwinism as proposed by Edelman et al., but differs substantially in its implementation. TNGS puts much more emphasis on selection of the fittest from a large but invariant repertoire of neural populations, not on replicating the Darwinian principles of heredity and mutation. The model is thus to be seen as modelling the principle as it is phrased verbally, not the formal circuits proposed alongside with this formulation.

The third simulation is an incomplete study with a more positive objective. Rather than just to pick holes in other people's arguments, it tries to explore the possible integrated and non-modularised functions that correlated activity in a neural module can have. It has, however, not been pursued to the point where it would have produced interesting results.

5.2 Model

The model investigates the question of the possible functional role of 'value systems' in a deliberately minimal toy-like set-up. It does not aim to model actual brain structures. It just serves to illustrate a conceptual argument of what correlated activity can mean *in principle* and what follows from the core assumptions underlying value system architectures *in principle*, if no additional assumptions are made.

A circular two-wheeled agent of four units diameter is evolved, in the first stage of the simulation, to seek the light (phototaxis) and, at the same time to generate a motor signal that correlates with its behavioural success (in analogy to a value system). Behavioural success is measured as relative distance from the light source.

In the second experiment, the internally generated value signal evolved during the first experiment is used as reinforcement signal for continued evolution of behaviour. This continued evolution is an analogy of value-guided neural Darwinism as proposed in TNGS.

The agent is controlled by a CTRNN (see chapter 3 equation (3.2)) whose structure (i.e., the connectivity C and the number of hidden neurons) is partially evolved. Connections to input neurons or from output neurons are not permitted. Input neurons can project to output neurons and to hidden neurons, hidden neurons can project to other hidden neurons and to output neurons. The network has two input neurons and five output neurons (specification below) and can have varying numbers (0-5) of hidden neurons. In experiments where the value signal E is integrated into the network dynamics (section 5.3.4), the estimator neuron changes status to become another interneuron. In some experiments, parts of the network structure and parameters were fixed and exempted from continued evolution at a certain stage (specified further down). The existence or non-existence of hidden neurons and neuronal connections is determined by the step functions $x > 0.7$ and $x > 0.6$ respectively.

The GA is the standard GA specified in chapter 3 section 3.3 (vector mutation with $r = 0.7$, usually 2000 generations). Parameter ranges are $w_{ij} \in [-8,8]$, $\theta_i \in [-3,3]$ and $\tau_i \in [16,516]$.

The agent has two light sensors $S_{L,R}$ with an angle of acceptance of 180°, which are oriented $+60^\circ$ and -60° from the direction in which the agent heads. The sensor orientation is subject to uniform directional noise $\in [-2.5^\circ, 2.5^\circ]$. Their activation is fed into input neurons by $I_{Si}(t) = S_G \cdot S_{L,R}(t)$ with the evolved $S_G \in [0.1,50]$ and $S_{L,R}(t) = 1$, if the light is within the sensory range at time t and $S_{L,R}(t) = 0$ otherwise. The *binary activation of light sensors makes the fitness estimation non-trivial*, as there is no direct signal present in the sensory inputs that represent distance from

the light source (e.g., light intensity). In order to generate a motor signal that corresponds to behavioural success, an active perceptual strategy has to be evolved.

The motor velocities are set instantaneously at any time t by $v_{L,R}(t) = M_G(\sigma(a_{L1,R1}(t)) - \sigma(a_{L2,R2}(t)) + \varepsilon$ where M_G is evolved $\in [0.1, 50]$ and $a_{L1,L2,R1,R2}$ is the activity of the four motor neurons generating the velocity. $\varepsilon \in [0, 0.2]$ is uniform motor noise. A fifth output neuron n_{M5} generates the performance estimate $E(t) = \sigma(a_{M5}(t))$ which, during the first experiment, is evolved to represent the present distance to the light source relative to the starting distance to the light source (fitness function equation (5.3)).

In every evaluation, the agent is presented with a sequence of 4-6 light sources that are placed at a random angle and distance $d \in [40; 120]$ from the agent. Evaluation trials last $T \in [3000, 4000]$ time steps. They are preceded by $T' \in [20, 120]$ simulation time steps without light or fitness evaluation, to prevent the initial building up of activity in the estimator neuron from following a standardised performance curve. Each light is presented for a random time period $t_i \in [\frac{T}{5} - 100, \frac{T}{5} + 500]$ time steps. The network and the environment are simulated with $h = 1$.

The fitness $F(i)$ of an individual i is given by

$$F(i) = F_D(i) \cdot F_E(i) + \varepsilon F_D(i) \tag{5.1}$$

where $F_D(i)$ rates the phototactic behaviour and $F_E(i)$ rates the fitness prediction. The second term ($\varepsilon = 0.001$) is included to bootstrap the evolutionary process by minimally rewarding light-seeking behaviour over no sensible behaviour. The co-evolution of light seeking and estimation of performance using the product of both terms is difficult for evolutionary search to generate from scratch. I chose to use the product of the two terms rather than a weighted sum because it was very difficult to evolve estimation behaviour and any relaxation on the selection pressure for evolving a good estimate and light seeking at the same time immediately led to the evolution of a trivial standardised activation curve in n_{M5} that correlated just well enough with behavioural performance.

$F_D(i)$ is given by

$$F_D(i) = \frac{1 - P^2}{T} \sum_0^T max\left(0, 1 - \frac{d(t)}{d(t_0)}\right) \tag{5.2}$$

with $P = \frac{0.125}{T} \sum_0^T \frac{v_L(t) - v_R(t)}{M_G}$ included to discourage turning. $d(t)$ is the distance between robot and light at time t and t_0 is the time of the last displacement of the light source.

It was difficult to find a formulation for the fitness estimate F_E and it has undergone a number of necessary refinements before arriving at the following formulation

$$F_E(i) = \sqrt{\max\left(0, \frac{e(\bar{d}, d) - e(E, d)}{e(\bar{d}, d)}\right) \cdot \max\left(0, \frac{e(0, \dot{d}) - e(\dot{E}, \dot{d})}{e(0, \dot{d})}\right)} \tag{5.3}$$

with $e(x, y)$ the sum of squared error $e(x, y) = \sum_0^T (x(t) - y(t))^2$. $\bar{d}$ is the average of $d(t)$ during each trial. $\dot{d}(t)$ and $\dot{E}(t)$ are the derivatives of $d(t)$ and $E(t)$ averaged over a sliding time window $w = 250$ time steps (sum borders for $e(x, y)$ have to be adjusted accordingly). I chose to evaluate both the absolute error in predicting fitness and the error of following the changes in distance in order to avoid that networks evolved standardised curves of neural activity that only vaguely corresponded to the real changes in distance. Similarly, rewarding only estimates that are better than

average estimates forced artificial evolution to generate solutions other than constant or constantly increasing signals.

Fitness evaluation is exponential across $n = 6$ trials as defined in section 3.3, equation 3.4.

During the second experiment, the fitness function F_i in equation 5.1 is substituted for the value signal (distance estimates) E, such that

$$F'(i) = \sum_{0}^{T} E(t) \tag{5.4}$$

For the third (incomplete) experiment, I used the original fitness function defined in equation (5.1). The difference is that the neuron that is evolved to generate the value signal was promoted from motor neuron to hidden neuron and could thus take a functional role in the constitution of neural dynamics and the performance of light-seeking behaviour.

5.3 Results

I first present the results from the co-evolution of light-seeking and fitness estimation behaviour (subsection 5.3.1). In this section, I do not talk about statistical properties of the evolution but analyse the behaviour of one agent that has evolved a very effective yet simple strategy which helps to illustrate the conceptual argument. In subsection 5.3.2, I present results from the second part of the experiment, i.e., the continued evolution of behaviour substituting the external fitness criterion for the internal value signal as an analogy of value-based learning. Subsection 5.3.3 adds some results on co-evolvability of performance estimation and phototaxis, before subsection 5.3.4 gives an informal description of some preliminary results from an extended simulation in which a value signal was integrated into the network architecture such that its function for adaptive behaviour is a result of the evolutionary process rather than built-in, as in subsection 5.3.2.

5.3.1 Co-evolution of Light-Seeking and Fitness Estimation Behaviour

The network controller evolved to control the two-wheeled simulated agent is extremely simple, but astonishingly good at estimating how close the agent is to a light source, despite the minimal sensory endowment (two light sensors generating on-off signals) and the consequent ambiguity in the sensory space (i.e., any sensory pattern could occur at any distance from the light source). Even though these structures were at the GA's disposal (as explained in section 5.2, the evolved controller has no hidden neurons, recurrent connections or slow time constants. Therefore, its behaviour hardly relies on internal state and its complexity is minimal, even within the already restricted range of possibilities.

As a consequence of the absence of recurrent connections and hidden neurons, the neural sub-structure that generates the value signal is structurally isolated from the rest of the network dynamics, apart from being fed by the same input neurons (see encircled group of three neurons in 5.3). This strict modularisation of evaluative and behaviour generating neural substrates had not been built-in, but the fact that it resulted from the evolutionary process makes the analogy with value system architectures even stronger.

When analysing what this 'value system' does, I found that in the absence of light, or if the network receives input only on its right light sensor ($S_R = 1, S_L = 0$), it estimates $E \approx 0$. If light is

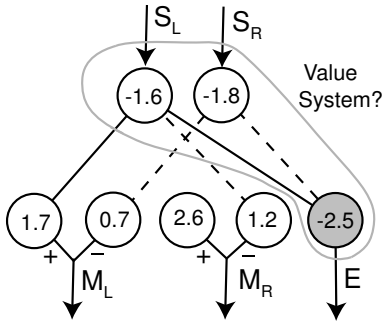

Figure 5.3: The controller of the agent that seeks light and estimates its distance from the light. (θ in neurons, dotted lines interneural inhibition, solid lines interneural excitation.)

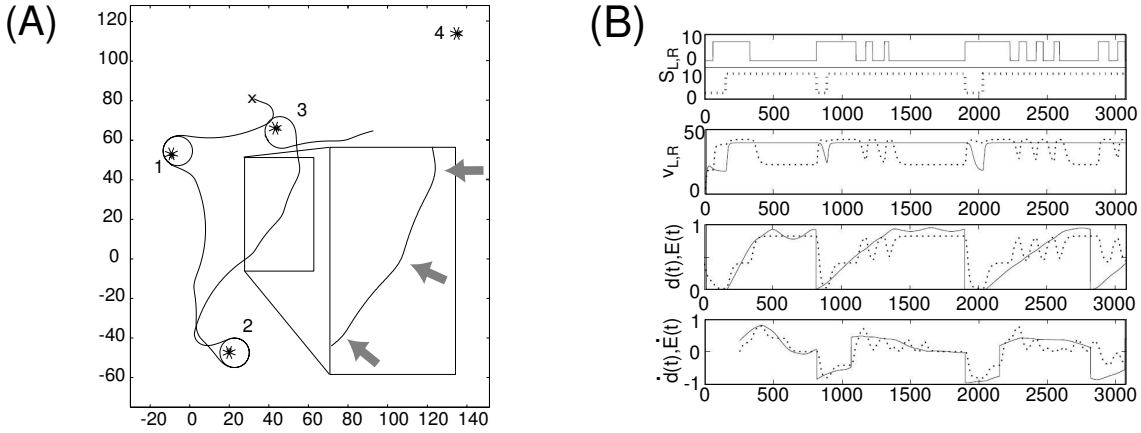

Figure 5.4: (A) Successful light seeking trajectory for four presentations of light sources. Arrows indicate the punctuated turns during $t = 2200 - 2700$ (see text). (B) The evolution of different variables over time in the same trial (Top to bottom: $S_{L,R}$, $v_{L,R}$, $d(t)$ vs. $E(t)$, $\dot{d}(t)$ vs. $\dot{E}(t)$).

perceived with both sensors, it estimates $E \approx 0.5$, and if the network receives input only in its left light sensor ($S_R = 0, S_L = 1$), the estimate reaches its maximum of $E \approx 0.8$. The judgment criteria of this value system can thus be described as 'seeing on the left eye is good, seeing on the right eye or not at all is bad'. Taken by themselves, these rules do not make sense.

Nevertheless, as we can see in figure 5.4 (B) (bottom two plots), both $E(t)$ and $\dot{E}(t)$ (dotted lines) follow with surprising accuracy the actual values $d(t)$ and $\dot{d}(t)$ (solid lines), particularly if we remember the poor sensory endowment of the agent.

In order to understand how this accuracy in estimating the performance is achieved, it is necessary to take into consideration the agent's light seeking strategy (figure 5.4 (A) and (B)). The agent's phototactic behaviour is realised by the network minus the estimator neuron. In the absence of sensory stimulation, the agent slowly drives forward, slightly turning to the right. Thereby, it draws a circle that will eventually make the light source appear in its visual field, entering from the right. If $S_R = 1$ and $S_L = 0$, the 'brake' on the left motor M_L is released and induces a sharper turn to the right. This means that the light eventually crosses into the centre of the visual field of the agent, i.e., $S_R = 1$ and $S_L = 1$, which triggers the agent to release the 'brakes' on both wheels

and drive almost straight, only slightly drifting to the right. This right drift in the near straight approach behaviour means that the light source repeatedly disappears from the right sensor's angle of acceptance ($S_R = 0$ and $S_L = 1$), which induces a sharp turn to the left that brings the light source back into the range of the right light sensor ($S_R = 1$ and $S_L = 1$). Once the light source is reached, this sharp turning to the left results in circling anti-clockwise around the light source, as this ongoing sharp turning to the left does not bring the light source back into the sensory range of S_R. In combination, these phases lead to the following sequence of behaviour during the approach of a single light source: 1.) A scanning turn to the right, until $S_L = S_R = 1$. 2.) A quick approach of the light from the right side, bringing the light source in and out the sensory range of S_R (cf. the rhythmically occurring drops of sensory and motor activity in figure 5.4 (B)). This strategy results in the chaining of nearly straight path segments in the approach trajectory, separated by punctual left turns (arrows in figure 5.4 (A)). 3.) counter clockwise rotation around the light source during which the light source is perceived with the left sensor only.

Knowing about this light seeking strategy, it is much easier to understand how the 'value system' achieves a correct estimation of the distance: the approach behaviour only starts when the light is in range of the left light sensor, and this sensor remains activated from then on, which explains the positive response to left sensor activation $S_L = 1$. $S_L = 0$, on the other hand, implies that the light has not yet been located, which only happens in the beginning of the trials if the agent is far away from the light source, hence $E \approx 0$. The right light sensor is activated during the approach trajectory, but not once the light source has been reached. Therefore, it mildly inhibits n_{M5} which results in $E \approx 0.5$ when $S_L = S_R = 1$. An additional level of accuracy during approach behaviour is achieved by keeping the light source at the boundary of the right sensor's sensory range by approaching the light at an angle from the right: the closer the agent is to the light source, the larger the angular correction necessary to bring the light source back into its sensory range and, therefore, the longer the intermittence in right sensor stimulation (see figure 5.4 (A), little arrows, and (B), oscillations in sensory input and estimation). This implies that, on average, the fitness estimate is higher the closer the agent is to the light source, because the right sensor, which mildly inhibits the performance estimate, is switched off for longer intervals. When the agent has reached the light source and cycles around it, $S_R = 0$ and $S_L = 1$, and the value system produces its maximum estimate $E \approx 0.8$, expressing that the light source has been reached.

Even if this phase of the simulation had mainly been intended to provide the basis for the second part of the simulation, i.e., an agent that generates a certain behaviour and a signal that represents behavioural success (value signal), it demonstrates an important theoretical point in itself: a value signal that correlates to behavioural success, even if it is generated by a neural structure that is modularised and not linked to the systems that generate motor behaviour is not necessarily disembodied and explicable outside the sensorimotor context. This is an interesting theoretical insight with respect to the question of neural correlates of behaviour: a neural cell assembly that is identified to generate a neural signal that correlates with behavioural success is not necessarily solely responsible for generating this signal, even if the neural structure is fully separated from the structures that generate sensorimotor behaviour. The external closure of the sensorimotor loop can establish a link that is absent in neural connectivity.

Another event worth discussing in the trial depicted in figure 5.4 (A) and (B) occurs after the

last displacement of the light source ($t > 2800$): as the displacement happens to bring the light source into the left visual field of the agent, it immediately enters the oscillating approach mode and its estimate therefore poorly corresponds to the actual distance measure which drops to 0. This dissonance can be seen as a possibly inevitable error due to the minimalism of the sensory equipment of the agent. However, putting oneself 'in the agent's shoes', it could also be interpreted as the superiority of the evolved estimator over the distance measure as a measure of performance: The comparably high output expresses the agent's justified optimism to be at the light source soon, which is not reflected in the distance fitness measure $F_D(i)$, which evaluates distance independent from sensory state, orientation of the agent and what they imply for behavioural success. Such discrepancies between meaningful judgment signals generated by the agent and *a priori* specified performance measures were one of the key difficulties in designing the experiments. Even with the highly refined and complex fitness measure F_E (equation (5.3)), sometimes, 'good' solutions in terms of how behaviour estimation corresponds to progress in terms of meaningful actions were replaced with less sophisticated ones that better corresponded to Euclidean distance.

5.3.2 A Caricature of 'Value-Guided Learning'

As explained in section 5.1, in explaining the mechanisms of life-time learning through neuronal group selection (e.g., Edelman, 1989; Sporns & Edelman, 1993; Edelman, 1987) the proponents mention only the following components

1. A neural assembly whose activation correlates to saliency of events (value system).
2. Neural selection based on Darwinian principles that is guided by the activity in the value system.
3. The possibility for value system learning supervised by higher order value systems.

The second simulation model whose results are presented in this section only investigates the first two of these three components. The third point is discussed in the discussion section 5.4. The model simulates the logical consequences of the mentioned principles if implemented without further support structures. In this study, the evolution of the robot controller is seen as the analogue of ontogenetic neural Darwinism as proposed in TNGS. As mentioned above, the ER model is inspired by TNGS but differs in the implementation of neural Darwinism. It models the general principle of separating pre-specified meaning generation and plastic behaviour generation, which is much more commonly adopted than just in TNGS.

The GA is seeded with a population of the successful individual discussed in the previous subsection 5.3.1. The only parameters that evolve in this experiment are the strengths of the three synaptic connections from sensors to motors (behaviour generating sub-system; cf. figure 5.3). The fitness measure F is substituted for the performance estimate $E(t)$ (equation (5.4)). It is important to notice that in this set-up, the value system does not evolve, it just guides the evolutionary change of the synaptic weights to reinforce whatever behaviour leads to a high performance estimate $E(t)$. Value systems are the proposed neural structures to guide ontogenetic adaptation. But can such mechanisms work if the value system is properly embodied?

Figure 5.5 (A) illustrates how with the embodied value system evolved in the first phase of the experiment (whose judgment relies on an established sensorimotor strategy), 'value-guided

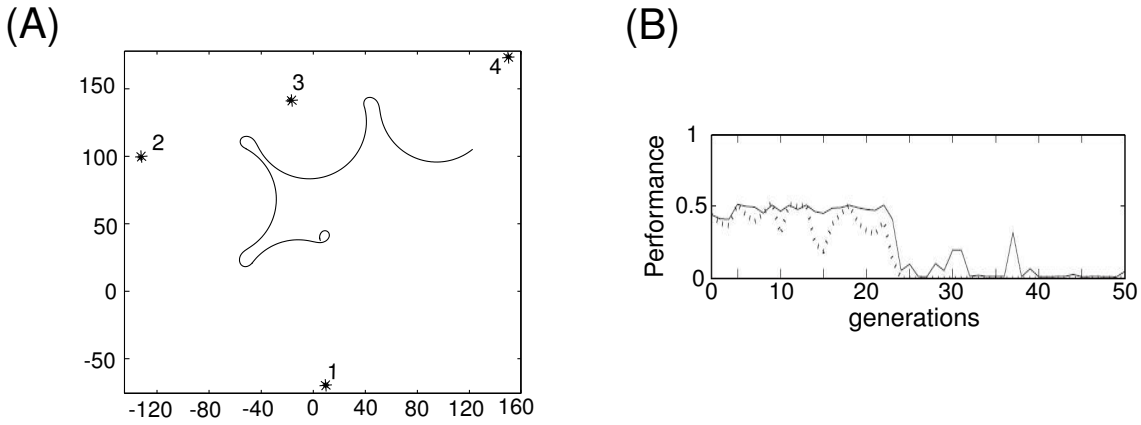

Figure 5.5: (A) Light-avoiding trajectory of an agent after 50 generations of 'value-guided learning'. (B) The degeneration of light seeking performance F_D (solid line) and estimation performance F_E (dotted line) over generations (learning) for the same experiment.

learning' results in a complete deterioration of phototactic behaviour within 50 generations (figure 5.5 (B)). Behaviour is altered to driving around the light source in large anti-clockwise circles, not approaching at all, which results in a deterioration in both components $F_D(i)$ and $F_E(i)$ of the fitness evaluation, as value judgment is maximally positive.

This deterioration is a consequence of the plastic sensorimotor context in which the value system generates its judgments. In a variable sensorimotor context, what the 'value system' rewards is simply activation of the left light sensor but not the right, as discussed in the beginning of section 5.3.1. The fact that this judgment of sensory activation means good light seeking behaviour during embodied interaction is a contribution of the sensorimotor context, and this meaning is removed if the system is functionally separated from the sensorimotor context. The gradual change of behaviour results in what I termed *semantic drift*: activity in the value system causes a change in behaviour, which in turn causes a change of 'meaning' of the activity of the value system, which causes a change in behaviour, and so on. The system described above, in isolation, rewards activity of the left sensor and punishes activity of the right. If this semantic contribution of the sensorimotor couplings to the function of the value system is gradually modified, the agent ends up avoiding the light source in a large circle, because this is the behaviour that optimises value system output – even if it is not phototaxis.

This deterioration of performance is hardly surprising, given the structure of the value system and its functioning as discussed in subsection 5.3.1. It demonstrates an important theoretical point, though: it shows that value system architectures based on only the two principles mentioned above are not guaranteed to work. Further modelling assumptions have to be taken onboard. Most crucially, it has to be ensured that *a value system's judgment works despite the plasticity of the behaviour that it supervises*. A value system has to be disembodied in the sense that it is not subject to changes in meaning in the presence of reciprocal causal links, feedback loops and semantic drift of local structures. I return to this point in the discussion.

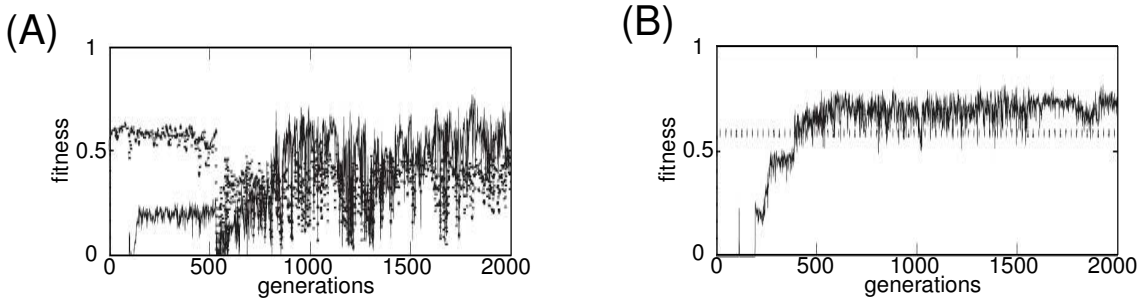

Figure 5.6: Performance profile across evolution for (A) co-evolution of evaluation and light seeking and (B) evolution of fitness estimation given a fixed phototactic behaviour. F_E (solid) and F_D (dotted).

5.3.3 Evolvability

Comparing the agents evolved to estimate value and seek lights to agents evolved to achieve just phototaxis (i.e., $F(i) = F_D(i)$), I found that the light seeking behaviour in agents that are evolved to estimate their performance at the same time is clearly suboptimal.

I hypothesised that, in order to be able to judge the distance to the light-source, efficient phototaxis had to be somewhat traded for activity that enables the agents to perform judgments about the distance to the light. In order to test this hypothesis, I seeded evolution with agents successfully evolved to perform phototaxis ($F(i) = F_D(i)$) and evolved estimation behaviour on top (fix sensorimotor behaviour and evolve with $F(i) = F_E(i)$) and compared evolvability with the simulation described in section 5.3.1 (average across 10 evolutionary runs).

Had my hypothesis been true, the latter would have evolved better, because light seeking behaviour could have been modified by evolutionary search to allow better estimation of performance. However, I found that both F_D and F_E were on average higher in the agents with fixed sensorimotor behaviour (see figure 5.6 (B)). If good light seeking and good value estimation are possible at a time, why does the evolutionary search not find this solution? Taking a closer look at the co-evolutionary scenario in figure 5.6 (A), it becomes clear that the coevolutionary scenario is much noisier and good solutions repeatedly deteriorate. In the presented set-up, a good estimation of the agent's performance is very sensitive to behavioural modification and thus evolves to higher levels if there is no interference between the continued variation of both. This sensitivity in the given task contributes to explaining why 'value guided learning' leads to such a rapid and devastating decay of behaviour: the noise sensitivity of value estimation accelerates semantic drift.

Out of the evolutionary runs that evolved estimation behaviour on top of successful phototactic behaviour, one produced an agent whose light seeking strategy appears to make distance estimation truly impossible. This suggests that there is at least some need for sensorimotor behaviour to accommodate judgment, even if this accommodation is not necessarily reflected in inferior performance.

5.3.4 The Evolution of Value System Function

The purpose of the simulation model presented so far was primarily a criticism of the arguments of underlying approaches that presume a functional and structural division between meaning gen-

erating and behaviour generating sub-systems in the central nervous system. This objective is decompository and does in no ways exploit the generative capacity of ER simulations, i.e., to produce hypotheses and descriptive concepts. In the experiments presented so far, the value signals have no effect on the state of the neural controller or the environment other than the selection of synaptic weight it has been functionally assigned in the second experiment (subsection 5.3.2). There was no real possibility for it to serve any other adaptive function.

The reported correlation between neural activity and salient events in the environment is still an interesting one. This subsection presents the preliminary results from a model that aims to explore the functional role that such 'value signals' can play in constituting adaptive behaviour with minimal prior assumptions. In order to investigate this question, I decided to evolve a neural controller that a) exhibits phototaxis and generates neural activity that correlates anti-proportionally with the distance to the light (as specified in the original fitness function equation (5.1)) and b) to integrate the neuron that is evolved to generate a value signal into the network dynamics as an additional hidden neuron. Therefore, a value signal is part of the evolved neuro-controller and artificial evolution can use it as a dynamic building block for solving the task.

The most common structure I found in these networks is an excitatory self-connection in the estimator neuron that improves the estimation performance, but not phototaxis. In some of the networks that realise the same strategy described in section 5.3.1, light seeking performance crucially depends on the activity of the estimator neuron in that performance breaks down if the neuron is artificially lesioned or its activity is clamped. Looking at the functional role it bears, it becomes obvious, though, that it only serves to inhibit the right motor because its activity is roughly in inverse correlation with the activity of the right sensor, and thereby takes part in inducing left turns if the light goes out of the right visual field. Its function is, therefore, to simply relay and invert the right sensory signal.

The present simple phototactic behaviour is probably too simple for a value signal to bear any interesting functional role. In an extended version of the present experiment (work in progress), I evolve agents on the simple associative learning task presented in (Tuci et al., 2002) in order to further explore the possible functional properties for value like signals that are more open ended. The experiments presented here already illustrate two different possible such functions: such correlated activity can be simply epiphenomenal and without function, as in the model evolved in section 5.3.1. It can be plainly accidental in a circuit that mechanically regulates certain behaviours, as in the behaviour just described. As a further theoretical point, it is important to note that in both of these cases, the signal that correlates to salience of events does not *internally represent* success in any form. Even in their simple forms, these results can thus be seen as a proof of concept that correlation of neural activity does not justify the reduction of function to a localised structure.

5.4 Discussion

Value system architectures, as many related architectures proposed, presume an informationally encapsulated rigid structure to provide a meaningful signal for an otherwise meaningless process. Findings about brain areas whose activity correlates with salient events in the environment are interpreted as evidence for the existence of such value systems in the nervous system. The present simulation models show that this reasoning is not stringent: in the first experiment, it is shown

that even a modularised brain area generating a value signal that is not directly connected to the behaviour generating neural subsystems can depend on sensorimotor dynamics through indirect linking via the agent-environment interaction. In the second experiment it is shown how, as a consequence of this indirect link, the previously intact value judgment can break down as a consequence of the behaviour modification it itself induces, as gradual change of the behavioural context induces a gradual change in judgment capacity. This phenomenon, which is a direct consequence of the existence of reciprocal causal links between value system and behaviour generating systems, is what I term 'semantic drift'.

In the simulated agent, localised neural activity could be observed to correlate with fitness. It turned out, however, that the reduction of function to mechanism was not justified and that the structure instead relied on embodied dynamics to perform its judgments. It is important to see the severe consequences of embodiment and functional integration of the 'value system'. Such an integration undermines the very concept of the value system as a top-down modulator of behaviour - which is nothing but to say that you cannot simply presume a complex non-linear dynamical system to act (approximately) like a linear system.

In this sense, the present simulation can be seen as an illustration of the problems associated with 'hybrid' architectures that feature functional and structural separation: "If a full-blown ghost in the machine has difficulties dealing with the variability of the external world, why would a vestigial ghost in the machine not face the same difficulties dealing with the variability of its bodily environment?" (Rohde & Di Paolo, 2006c).

I now want to return to the third principle mentioned in section 5.3.2, i.e., the possibility for value system learning. If value system function can be maintained in the presence of environmental or bodily variation, neural Darwinism can work. Indeed, the reported evidence about correlation between brain activity in certain neural modules and salient events in the environment shows that such mechanisms to keep up this correspondance exist, even if the existence of this correlation does not *a priori* explain anything about its function. It has to be asked, however, if explaining the mechanisms to maintain the generation of a meaningful value signal is not ultimately much more difficult than explaining the adaptation guided by this value signal.

Sporns and Edelman's (1993) conjecture that "different value systems interact, or that hierarchies of specificity might exist" (Sporns & Edelman, 1993, p. 969) seems to suggest that the maintenance and adaptation of value systems should also follow the principles of value-guided neural Darwinism. It is not clear to me how the mere recursive application of value-guided learning circuits would not lead to a *regressus ad infinitum* or, otherwise, require a magic master-value system to end this regress. Recent work by Edelman et al. (e.g., Krichmar & Edelman, 2002), as well as by other groups (e.g., Doya, 2002), however, appears to break with the idea of neural Darwinism as fundamental principle of ontogenetic adaptation. They extend the proposed framework to include other kinds of neural plasticity and meta-modulation, proposing different kinds of adaptive circuits for different kinds of modulatory sub-systems. These models are informed by recent neuroscientific evidence and are conceptually much more complex than the simple neural Darwinism modelled in this chapter. These extensions appear to confirm Rutkowska's assumption that "[increased] flexibility requires some more general purpose style of value" (Rutkowska, 1997, p. 292) than a value module could provide.

It has to be stressed that the point illustrated is a logical criticism - it is not to say that the proposed circuits could not exist in practice, as part of a more complex system, relying on additional premises that had not been made explicit in the original proposal and thus not built into my caricature model. Existence proofs, even though they teach us to be careful not to presuppose a functional modularity, do not exclude the empirical possibility of such structures. Maybe there are "simple criteria of saliency and adaptiveness" (Sporns & Edelman, 1993, p. 969) that can *a priori* specify what will be good and what will be bad *a posteriori* - but this will have to be proven empirically. Maybe, value system functionality can be kept intact by mechanisms of value system learning - but it has to be shown and argued how that would happen rather than to just postulate such mechanisms. Probably, the identified neural structures whose activity correlates to salient events in the environment play a fundamental role in value-appraisal and adaptation - but reducing value generation to these structures seems a category mistake in that it confuses mechanism and behaviour and cannot be justified on the basis of correlation alone. Surely, there are possibilities to make these kinds of circuits work if the reciprocal causal links are cut. There are robotic artefacts with a limited behavioural domain (Verschure et al., 1995) that successfully implement the adaptive circuits proposed as part of TNGS. But in order to be convincing as a biological theory of general adaptivity, it would be necessary to specify how such rigidly wired value systems would be realised in a living organism that is in constant material flux.

To cut a long story short, the point of this model is not to discourage the scientific study of the described neural structures or to discourage the use of value system models if the conceptual limitations are made explicit - the point is just to be wary to not get overexcited by discovering correlation and subscribe to a partially Cartesian and representationalist story only because it is more spectacular. By postulating pre-specified value systems without explaining how they work, the explanatory burden "[b]uck [is passed] to evolution" (Rutkowska, 1997, p. 292) and the real question of why something matters to the organism, in the sense outlined above for the enactive approach, is not addressed.

A pre-coded approach will always be functionally limited in its adaptive capacities. This is not to say that the adaptation mechanisms found in living organisms can do anything - but their limitations are material and physical, not functional. As we argue in (Di Paolo et al., 2008a), the most striking examples of value changes, which can shatter the functionality of established relations, are illness and other perturbations to the body (distortion or impairment) such as we find them in PS (cf. chapter 3 section 3.4), "[o]r consider a patient who, during the course of a disease, is subjected to increasing dosages of a pharmaceutical agent, with the result that he not only survives dosages of the drug that would be fatal to the average human being, but also that his metabolism relies on the medicine in a way that deprivation would cause his death" (Di Paolo et al., 2008a). It is really unclear how a pre-coded value function could keep up with these changes in significance that result from the dynamical re-organisation of the organism itself.

There are a lot of open research questions concerning the origins of value and the structures that realise life-time adaptation. How can these questions be addressed from a modelling perspective without stepping into a reductionist trap? One avenue already mentioned is the work in progress on evolving value system function by posing a task that requires lifetime adaptation and to evolve a value system like structure whose activity correlates with behavioural success but

to not specify the link between the two. This approach serves to investigate possible functional roles of neural activity that correlates with behavioural success in an unbiased way. It holds the potential to generate intuitions and hypotheses about the origins of modularity and the integrated function of generating and modulating behaviour. This approach can be seen as a merger of the simulation models presented in this section and the work on evolving life-time adaptivity in fixed weight neural controllers (Yamauchi & Beer, 1994; Tuci et al., 2002; Izquierdo-Torres & Harvey, 2007) and focuses more on the question of localisation of function than on the origin of values.

Another approach that is less related to the simulation model presented in this chapter is to simulate or artificially create value-generating processes (or something close) in a minimally biased way. Using ER simulation models for this kind of model is a bit tricky, because ER simulation models are teleonomical through the external specification of the fitness criterion and thus the purpose and function of behaviour. Some simulation models that aim at reducing this bias to the minimum or at making it as implicit as possible have been attempted (e.g., Di Paolo, 2000b; Di Paolo & Iizuka, 2008). Other ALife simulation models that do not use evolutionary optimisation search to specify the function but instead start from the mechanical or chemical level tackle this difficult question from the opposite direction (Varela et al., 1974; Ikegami & Suzuki, 2008) and both approaches together can, possibly, sandwich the theoretical principles of genuine value generation and how function and mechanism relate to a certain extent.

Concerning the methodological research question of the current dissertation, the simulation model presented in this chapter demonstrates the kind of contribution that such models can make to conceptual and philosophical debate: the model took the premises made explicit in value system architectures to its logical conclusion showing that from the postulated principles alone, adaptation cannot be guaranteed. The model also generated useful descriptive concepts (noticeably, the concept of 'semantic drift').

As concerns the impact that this model has had on the scientific community, it is not as evident as in the other four simulation models presented in this thesis. This shortcoming, in my opinion, is mainly due to the fact that this model has not properly published and publicised in its own right: it had only been published as a technical report (Rohde & Di Paolo, 2006c) and as part of a book chapter addressing a much broader topic (Di Paolo et al., 2008a). Also, even though the reactions amongst researchers that are strongly committed to the enactive paradigm to the results reported in this chapter was extremely positive, the original presentation of the results in (Rohde & Di Paolo, 2006c) was too biased and antagonistic to be perceived and acknowledged by the relevant community (i.e., 'on the fence' researchers, as Inman Harvey puts it). Given the nature of paradigmatic debate, political and presentational issues are much more important for this kind of simulation model than for more science oriented simulation models. In order to exploit the demonstrative potential of the simulation model presented in this chapter, it will be important to publish and promote it; in order to make it more pleasant, it would be advisable to extend the work presented in section 5.3.4 in order to produce a positive message and outlook from the model as well, not just to criticise. This extension and publication of the simulation work on value system architectures will be pursued beyond the current dissertation, in order to not waste its merits.

In one way, using simulation models in order to add formal rigour into conceptual debate is very satisfactory, because they directly address the most essential, high level and far reaching

questions, such as the origin and nature of value and meaning in adaptive behaviour. On the other hand, these kind of simulation models produce less tangible results and do not directly relate to scientific practice, producing hypotheses and suggesting new experiments, as the model of motor synergies presented in chapter 4 does. Both modelling approaches, therefore, have been shown to be valuable in the study of human cognition and behaviour in their own scopes and limits, as argued in chapter 3 section 3.3.

It seemed, however, slightly unsatisfactory to be able to address only either problems of scientific practice or conceptual problems about high level human cognition, not both at a time. The following two chapters (6 and 7) introduce simulation models of work in PS, a scientific approach that links both domains (as explained in chapter 3). This work shows how integrating ER simulation models into this already interdisciplinary approach can combine their potential to contribute to both conceptual debate and scientific practice at the same time. These simulation models lead up to the application of the interdisciplinary framework I developed in chapter 3 section 3.6 to the problem of sensorimotor adaptation to sensory delays and perceived simultaneity (chapters 8-12), for which I conducted both the modelling and the experimental work.

Chapter 6

ER in Dialogue with Perceptual Supplementation Research: Perceptual Crossing in a One-Dimensional World

---

In the preceding two chapters, I have presented Evolutionary Robotics simulation models applied to problems in the area of human motor control (chapter 4) and neuroscientific theory (chapter 5). The former has shown how ER's potential to generate hypotheses and proofs of concept can immediately resonate with hands-on scientific research. The latter shows the more abstract philosophical value of Evolutionary Robotics models, i.e., to point out implicitly held prior assumptions in a theory and illustrate logical consequences that are counter-intuitive or difficult to understand. As I developed in chapter 3 section 3.6, my aspiration was to combine the merits of both types of approach, i.e., the concreteness and 'meatiness' of the scientific ER modelling and the immediate applicability to questions of high level cognition of the philosophical model. Therefore, I developed the approach to model the results from minimal experimental work on perception and sensorimotor adaptation (Perceptual Supplementation, PS). This and the following chapter present the results from the first application of this modelling approach to data from experiments by the GSP in Compiègne to study the dynamics of human perceptual crossing in a one-dimensional and two-dimensional simulated environment respectively. These chapters lead up to the second part of the dissertation (chapters 8 - 12), in which I analyse in detail the problem of adaptation to sensory delays and experienced simultaneity, presenting the results from an interdisciplinary project, in which I implemented not just the ER simulation model for existing results from an experimental study, but developed both the model and the experiment at a time and in dialogue with each other.

I introduce the problem area and the results from the experiment on perceptual crossing in a one-dimensional simulated environment (Auvray et al., 2008) in the introductory section 6.1, before briefly giving details of the model in section 6.2. The modelling results are presented in section 6.3 and evaluated and discussed in section 6.4. These results have been previously presented (Di Paolo et al., 2008b, 2008a; Rohde & Di Paolo, 2006b).

6.1 Background: Perceptual Crossing Through Tactile Feedback

The question addressed by both my model and the experimental study on which it is based (Auvray et al., 2008) is about the role of global interaction dynamics in social interaction. Social interaction of two or more individuals is a process of reciprocal causality. Such processes can lead to the emergence of dynamical patterns and global invariants that cannot be explained or understood by studying its components in isolation (cf. chapter 2). This means that phenomena dynamically emerging from social interaction may not directly result from the individual capacities, intentions or actions of any of the partners. As De Jaegher (2007) argues in detail, the collective and global dynamics that characterise social interaction are frequently neglected when studying social cognition (in approaches such as 'theory of mind theory' or 'simulation theory'). Despite evidence to the contrary that suggests the importance of interaction dynamics in social processes (such as, for instance Kendon's findings "synchronisation between interaction partners happens only when their mutual expectations of each other are exceptionally well attuned in the interaction" (De Jaegher, 2007, p. 149); many more examples are given in the cited source), these traditional approaches focus on explaining individual capacities.

Auvray et al. (2008) have designed an experimental paradigm to study the dynamics of social interaction in a minimal simulated environment. Two blindfolded participants are placed in separate rooms and connected to a computer. The virtual world in which they meet to interact is one-dimensional and infinite, i.e., a tape that loops around (for technical and parameter details see (Auvray et al., 2008) and the model section 6.2, particularly figure 6.1). They can move left and right on the tape, and whenever they cross an object, they receive a tactile stimulation to their fingertip through a Braille display. Participants are asked to indicate with a mouse-click when they believe a stimulation is caused by another feeling sensing intentional entity. Participants are told that apart from the other participant, there is a fixed object (fixed lure, at different locations for each of the participants to avoid them clustering around a fixed object) and a mobile object (the attached lure that actually shadows the subject's movement at a fixed distance, which the participants do not know) in the environment. All of the entities have the same size in the simulated environment.

Therefore, the task is not only to distinguish moving and static objects, but to distinguish two entities that perform identical movement trajectories, only one of which is able to sense and respond to the encounter with the participant. In (Di Paolo et al., 2008b), we have compared this experimental set-up with Trevarthen's double-monitor experiments (Trevarthen, 1979), in which two months old babies were tested for their capacities to distinguish a live interaction with their mother, mediated through a screen, from the presentation of a previously recorded interaction via this screen. The difference between the mother's behaviour on the monitor between the two conditions is only whether she senses the child and reacts to its actions or not; her expressive behaviour, i.e., her motion, language, mimics, dynamics, voice etc. are identical between the two conditions. From the fact that babies get distressed and removed when presented with a previous recording, it is concluded that even two month olds are sensitive to social contingency. In the light of the outlined tension between holistic and individualistic views on social interaction, the question to be asked is: does such sensitivity imply internal cognitive recognition/detection mechanisms of whether an interaction is recorded or not on behalf of the infant or does the difference between the

two conditions emerge (partially) from the interaction, possibly involving much simpler mechanisms?

The results by Auvray et al. (2008) show that subjects are very successful at solving the task ($\approx 70\%$ correct responses), without previous training and in spite of the poverty of the sensory information provided by the minimal simulated environment (a simple bit sequence). Astonishing at first glance, the results are demystified after a simple analysis of the sensorimotor dynamics of the task and the strategy adopted by the participants to solve the task. Participants search for stimulation and engage in local rhythmic scanning movements with any entity encountered on the tape. This rhythmic activity can only result in stabilised interaction with the other, not with the attached lure. When making contact with the attached lure, the lure shadows the movements of the other participant, who, meanwhile still searches for stimulation and therefore does not act rhythmically and locally, like the participant would do in a real interaction. This analysis is backed by analysing the ratio of clicks per stimulation, which reveals that the probability of clicking after encountering the attached lure is equally high as the probability to click upon encountering the other. The 70 % accuracy results not from discriminatory capacity, but from the fact that the participants are much more frequently stimulated by the other than by the attached lure, due to the fact that interaction with the other is a stable attractor in the task given the search strategy, whereas interaction with the lure is not. Even though the distinction between the fixed lure and moving entities appears to be made on an individual level (less clicks for the fixed lure per stimulation), the distinction between the attached lure and the other appears to result mainly from the interaction dynamics.

6.2 Model

The ER simulation model of the task that I conducted was my first application of ER modelling to the area of minimal sensorimotor experiments with humans that I propose in chapter 3 section 3.6. Therefore, this model can be seen as a test-bed for the methodological framework I developed. By generating very simple artificial agents, I wanted to further clarify the sensorimotor dynamics of the task in tractable, noiseless, idealised and fully controllable settings. I was hoping to be able to confirm and enrich the insights already gained about the task and its solution, and to generate hypotheses for further data analysis and possibly further experiments. By illustrating how it is possible for the results to stem from the rich dynamics of the social process itself, rather than from individual capacities, I wanted to also bring support for interactionist stances in the study of social cognition. There have so far only been few ER models of social interaction (Di Paolo, 2000a; Iizuka & Ikegami, 2004; Quinn, 2001) to provide such important proofs of concept.

The virtual environment in the model is nearly identical to the one used in the empirical experiment. The length of the tape is 600 distance units and any entity on it (lure or participant) has a width of 4 units. One difference (due to a misunderstanding of the experimental results that had not been published at the time) is that, while participants were just administered a single tactile input at any point in time, the input to the CTRNN controllers (as defined in chapter 3, equation (3.2)) consists of four neighbouring receptive fields of width 1 unit. The network generates two motor signals $M_{L,R}$ for left and right movement, $S_G, M_G \in [1, 100]$.

The GA and evolutionary parameters follow generally the specifications outlined in section

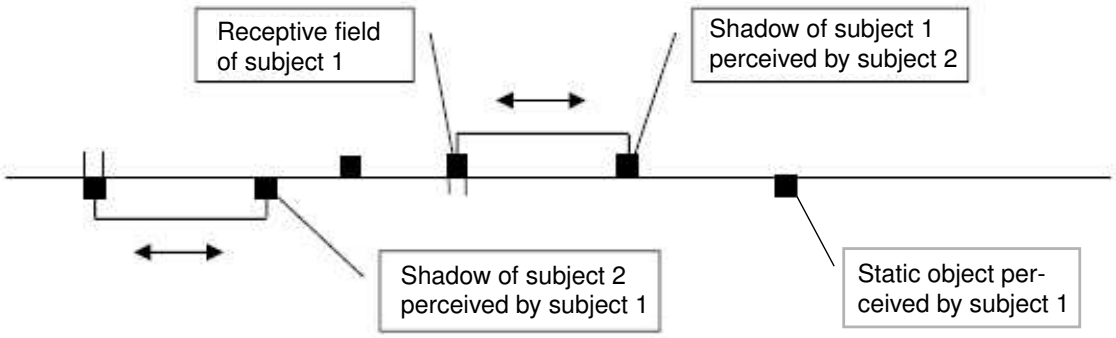

Figure 6.1: Schematic diagram of the one-dimensional environment in the perceptual crossing experiment.

3.3. The network structure, however, is modified and partially evolved. The two motor neurons are treated as hidden neurons in that the input neurons can connect to them directly and they can form recurrent connections with themselves or hidden neurons. There are up to 5 hidden neurons, and the network structure (i.e., existence of hidden neurons and synapses) is evolved using the step functions $x > 0.7$ (for connections) and $x > 0.6$ (for hidden neurons) respectively. Other parameter ranges are $\theta_i \in [-3,3]$, $\tau_i \in [20,3000]$ ms, and $w_{ji} \in [-8,8]$. In some runs, a sensory delay of 50 ms steps was applied. The trials lasted $T \in [8000,11000]$ time steps.

Agents are tested against clones of themselves using an exponentially weighted fitness average (equation 3.4)) over six trials, which constrains the space of solutions. The fitness criterion is the average distance $d(t)$ from the other across the trial

$$F = \frac{1}{T}\sum_{0}^{T} 1 - \frac{d(t)}{300} \tag{6.1}$$

The task is thus to locate the other agent and spend as much time as possible as close to each other as possible while not being trapped by static objects or shadow images. This is a slightly different task than that posed to the participants, who were not given any explicit encouragement to seek the other. They were only asked to indicate their perception of another sensing entity. As I found out in the later model of the two-dimensional version of the task (chapter 7), this modelling assumption biased the evolved behaviour to seek live interaction in a way that does not result naturally from the task. The reason for including this bias was to avoid the evolution of trivial but perfectly viable behaviour, i.e., to not seek interaction, which is not ruled out by the task as given to participants.

6.3 Results

First attempts to evolve agents to solve the described perceptual crossing task were unsuccessful. Evolutionary search got stuck in a local maximum to halt when crossing any object encountered on the tape, be it the partner, the fixed object or the attached lure of the other. Given the experimental set-up, this is a comparably successful strategy: If agents first encounter each other, or if one agent runs into its waiting partner, this strategy yields perfect fitness, and these are the majority of

possible cases. However, it is not the optimal behaviour, as in the remaining cases, the agents will not find each other at all, because they either stop on both of the fixed lures, at a maximum distance from each other, or in a configuration where one agent stops on the fixed lure and the other one on its attached lure, which will also yield low fitness. Also, it is not a very intelligent or adaptive solution and does not resemble any of the strategies adopted by human subjects, who keep actively exploring the environment, even after they have found the other, engaging in rhythmic interaction. Only when a 50ms sensory time delay between crossing an object on the tape and the agent's sensation was included into the model, active perceptual strategies evolved and the local fitness maximum of stopping when being stimulated could be overcome.

Where agents without delay evolved to simply stop, with the delay they evolved to engage in rhythmic interaction. This finding, which complies with the findings resulting from the two-dimensional model presented in the following chapter, shows that there is a relation between oscillating scanning movements and the delay in the evolved agents. This suggests that there may be a similar relation between the oscillatory strategies that most subjects adopt and the existent delays between sensation and reaction in humans. Even though it seems natural to us that subjects would adopt a strategy such as oscillatory scanning it is not *a priori* necessary, and it even seems like a waste of energy. There are many possible explanations for this behaviour, but we can derive the hypothesis from the model that sensory delays may play a role in generating scanning behaviour, and that subjects, possibly, repeatedly cross the partner because of reaction time delays, like the evolved agents do. This hypothesis could be tested in further experiments; it predicts that the amplitude of scanning oscillations is positively correlated with the amount of sensorimotor delay in a task where sensorimotor delay is varied between different conditions.

The trajectories the agents generate are similar to those generated by some human subjects (figure 6.2 (A); compare with the figures given in (Auvray et al., 2008)). This correspondence and the close match between the experiment and the model gives reason to hope that analysis of the synthetic results yields insights about the evolved solutions that generate quantitative hypotheses for the experimental data. This kind of data-driven prediction goes beyond the abstract proof of concept resulting from the model of motor synergies presented in chapter 4.

When taking a look at how behaviour evolved over generations, a consistent pattern is that avoidance of the attached lure evolved very quickly, while avoiding the fixed lure seemed to take a long time. These findings contradict the intuition that the easier task would be to recognize and avoid a static object, while distinguishing two entities that perform identical movements, only one of which responds to the perceptual encounter seems much harder. It could be argued that the apparent ease with which humans detect distinguish fixed objects from moving ones could be through detecting the invariant correlation between tactile and proprioceptive sensory input during active scanning, and that the artificial agents I evolved did not develop this strategy because they do not have proprioception. This is, however, only superficially true. The neuro-controllers evolved allow for recurrent feedback to be used. In the simple virtual environment modelled, re-afference of motor signal corresponds to proprioception and evolution could thus easily implement this strategy if it was advantageous.

A close look at the data from the simulation model suggests a different explanation. The strategy evolved in the artificial agents is simple: invert your movement direction if you sense

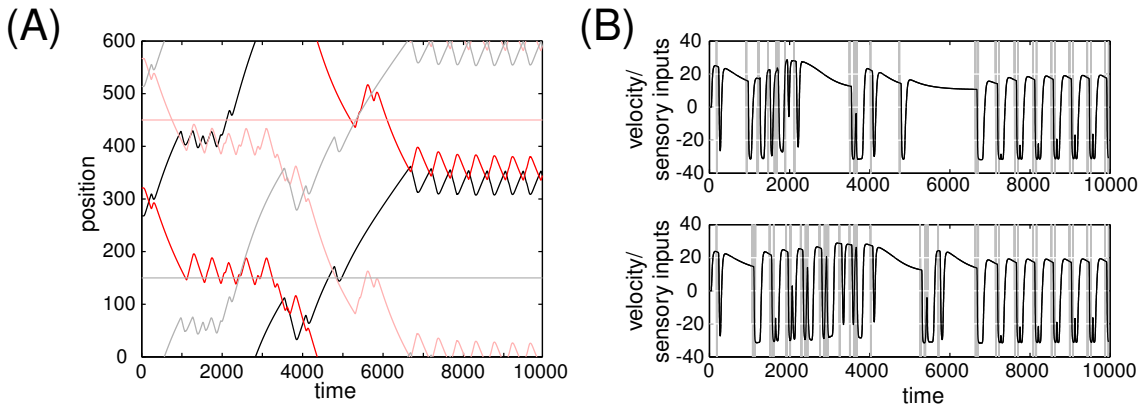

Figure 6.2: Results from evolved model. (A) A trial resulting in stabilised perceptual crossing with motor noise (position across time; Agent 1 black, agent 2 red; attached and fixed lures grey and pink). (B) Sensorimotor values for the behaviour depicted in (A). Agent 1 top, agent 2 bottom; velocity black, sensory inputs grey.

something, thereby crossing what you encountered again, and if sensation ceases in reverse crossing, go back to your original velocity, cross it again, invert, etc. Taking a close look at how sensation and motion evolve over time with this strategy, there is a striking similarity between the patterns observed during rhythmic coordinated mutual scanning (crossing) and rhythmic scanning of a fixed object (see figure 6.3 (A) and (B) bottom). This is because when both agents engage in this motion pattern, they will always meet in the middle of their return trajectory at the same location in the virtual space (see figure 6.3 (A), dotted line). This coordinated activity leads to sensations and motions changing over time in a way very similar to those that come about when investigating a fixed object (see figure 6.3 (B)).

So, what is the strategy employed by the agents in order to distinguish coordinated interaction and a fixed object? The duration of the stimulus upon crossing a fixed object lasts longer than when crossing a moving agent because the agent, even though it is the same size as the fixed object, moves in the opposite direction. The solution that the simulated agent adopts simply relies on integrating sensory stimulation over a longer time period, which yields a higher value for a static object, i.e., it is sensed as having a larger apparent size. Further support for this explanation comes from the fact that agents are quite easily tricked into making the wrong decision if the size of the static object is varied, i.e., a small object is mistaken for another agent and a larger agent is perceived as a fixed object.

It is interesting to note that the smaller perceived size in the case of perceptual crossing depends on encounters remaining in anti-phase oscillation, which is an *interactionally coordinated property* as defined in (De Jaegher, 2007). Thereby, the agents co-construct the appearance of the agents being of smaller size. The velocity inversion upon stimulation is tuned to this smaller perceived size such that the timing of the two perceptual crossings and the magnitude of the velocity inhibition they induce lead to a repeated coordinated oscillation around a fixed point of interaction. In turn, individuals respond to this emergent coordination by remaining in coordination with the apparently smaller object (see figure 6.3 (A)). In the case of the scanning of the fixed object,

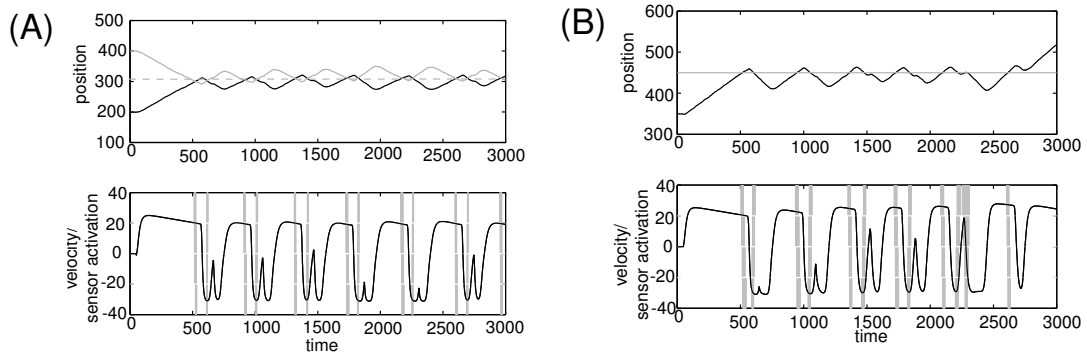

Figure 6.3: Trajectories and sensorimotor values of interaction with a fixed object and with the other (details). (A) Stabilised perceptual crossing between two agents (trajectories and sensorimotor values; dotted line location where perceptual crossing repeatedly takes place). (B) Scanning of a fixed object (trajectories and sensorimotor values). All diagrams include motor noise.

however, the longer sensation induces a longer return trajectory, temporally displacing the second crossing of the object, which means that the agent crosses incrementally further over the object with every perceptual crossing, disrupting the evolved stable interaction pattern which relies on timing of sensation and motion particular to perceptual crossing (see figure 6.3 (B)). The fact that the discrepancy between the two conditions amplifies over several crossings makes this strategy rather robust to motor noise. Disruptions of the required spatiotemporal interaction patterns can be cancelled out upon the next crossing.

6.4 Discussion

There are many viable solutions to the task, and it is rather unlikely that humans would use a strategy as the one just described, as it appears rather specific to the conditions under which the agents were evolved. Even though the trajectories look qualitatively similar, the algorithmic preciseness with which interaction is initiated and maintained is very unlikely to be found in the empirical data. But, as I argued in chapter 3, the point in modelling is not to recreate the original phenomenon but to identify invariant dynamical principles in an idealised and more tractable simulation.

In a similar way as the findings from the model of synergies as a principle in motor control presented in chapter 4, the present simulation model generates a number of proofs of concept that are valuable in an abstract sense. As argued in the introduction section 6.1, most of the research in social cognition is individual-centred. The modelling approach I took in this chapter does not just look at the individual capabilities, but also at phenomena that emerge during embodied and situated interaction. This broadened perspective uncovers relevant factors that are easily overlooked otherwise: A task that intuitively seems difficult, i.e., to distinguish two entities with identical movement characteristics (the partner and the shadow image), becomes almost trivial, if the effects emerging from the mutual search for each other are taken into consideration. This finding already results from the minimal empirical closed loop experiments by Auvray et al. (2008). However, they become, in a sense, even stronger through my simulation experiments because we can say *for sure* that there is nothing more complex going on in the artificial agents because the

network controllers evolved are extremely simple, too simple to do anything more sophisticated than what we analysed to be the strategy in section 6.3.

Also, the present simulation points out a different counter-intuitive state of affairs: Distinguishing a moving entity (the other agent) from a static one, which intuitively seems very easy, is indeed a non-trivial task, if the emergent effects of interaction, i.e., anti-phase coordination, are taken into consideration. In the original experiment, 32.7% of the stimulation were caused by the fixed object, as opposed to 15.2% caused by the attached lure. This suggests that participants may similarly find the intuitively easier task of avoiding the fixed object more difficult, even if this increased difficulty does not manifest in classification mistakes (as explained in the introduction section 6.1). There is evidence from both the model and the experiment that the distinction that arises mainly from interaction dynamics (which moving object is the other agent?) is more efficiently solved than the distinction that requires individual recognition capacities (Is the entity I am scanning the fixed object or the other?).

With this global view on the dynamics of perceptual crossing in the investigated set-up, these insights may seem almost trivial. However, had we started from the perspective of the individual and its conscious recognition capacities (such as 'theory of mind' approaches in social cognition), these findings would be mysterious - just as Trevarthen's (1979) results from the double monitor paradigm seem mysterious when focusing on the individual perspective, not on the interaction dynamics.[1]

However, due to the close match between the model and the experiment, I can go further than with the models presented in the previous chapters and generate quantitative hypotheses about the gathered data. Based on the findings on how integrated sensory stimulation time due to differences in perceived size of the other and the fixed object, I could hypothesise that, in the human experimental data, one of the predictors for this decision will be an apparently smaller object scanned. The researchers of the GSP favour a different explanation for this decision, i.e., "something that resists being spatially determined" (Auvray et al., 2008, p. 18), which is valid for the experimental data but not for the noiseless model. Interestingly, however, the experimental data *also* supports the explanation generated from the model, i.e., decreased stimulation time due to opposed movement is a good predictor for when participants click ('event E6' in (Auvray et al., 2008)). Therefore, the model generated a quantitative prediction about human behavioural data. To my knowledge, this was the first case in which an ER simulation model predicted human behaviour. Another quantitative prediction generated from the model has already been mentioned in the results section 6.3, i.e., that there would be a proportional relation between sensorimotor latencies and the variation in the magnitude of oscillatory scanning.

As it was the case for the modelling of motor synergies, I was very happy to find that the experimental researchers (and my later collaborators) acknowledged the significance of the contribution my model made. When publishing their results (long after conducting the study and the implementation of the present model), they wrote

> "In addition, the paradigm reported here can easily be implemented in evolutionary robotics computer interactions in order to address the issue of the sensitivity to percep-

[1] This logic also works the other way around: when Tom Fröse summarised Trevarthen's work in an 'Alergic' seminar in Jan. 2008, the audience, which mainly consisted of computational neuroscientists, biologists and people working in dynamical systems approaches was not in the least impressed or surprised by the baby's behaviour.

> tual interactions. It should be mentioned that Di Paolo and his colleagues (Di Paolo, Rohde, & Iizuka, 2008b) have already started a program of simulation research in that direction, based on the paradigm and methods reported in this paper. Their evolutionary robotics simulations showed similar results as the one reported in our study. Interestingly, and contrarily to any a priori prediction, Di Paolo and his colleagues found it easier to evolve agents that can distinguish between the avatar and mobile lure than agents that can distinguish between the avatar and fixed object. As a consequence, according to Di Paolo and his colleagues, in the case of social interactions, it is simply not necessary to evolve simulated agents with an individual contingency recognition strategy, given that the social process takes care by itself of inducing the individuals to produce the right behavior" (Auvray et al., 2008).

It is very reassuring that experimental researchers value the contribution that my simulation model makes to the understanding of the real-world phenomenon investigated enough to credit it in the scientific publication of their experimental work.

One important difference between the set-up investigated by Auvray et al. (2008) and Murray and Trevarthen's double TV monitor experiments (Trevarthen, 1979) is that in the double TV monitor experiments, the baby is only either confronted with its mother or with a recording of its mother, whereas in the experiments on perceptual crossing, the other participant and its attached lure are presented at the same time. It could be argued that the dynamic distinction emerging from the interaction dynamics in the perceptual crossing experiments is specific to the set-up because of the linkage between the attached lure and the other participant. As long as the other participant is still searching, the attached lure keeps moving away, shadowing the search trajectories and making stable interaction impossible, which is not the case in the double TV monitor experiments: infants could, in principle, adapt in one-sided interaction with the recording of their mother.

Our collaboration on modelling the dynamics of perceptual crossing (Di Paolo et al., 2008b) does not just present the simulation results presented in this chapter but also an extended model, implemented by Iizuka and Di Paolo (2007), which is, in this point, closer to Trevarthen's paradigm. This model shows that, in an equally simple simulation model as the one presented in this chapter, artificial agents evolve to distinguish between a recording and live interaction using a very simple, yet very effective action-perception strategy. Agents oscillate around each other, in anti-phase oscillation. If a previous interaction is replayed using identical starting positions, similar behaviour is observed initially, which is a case of one-sided interaction. However, the agents sporadically induce perturbations (jump sideways) into the apparent interaction to 'test' whether the interaction is live. What sounds like a complicated behaviour including elaborate planning and decision making, effectively is just a very simple sensorimotor coupling in which a reflex causes the breakdown of one-sided coordination. If the agent interacts with a recording, the breakdown is irrecoverable due to the lack of mutuality in the interaction, whereas in a genuine two-sided interaction, the other agent reacts to the perturbation induced and restores rhythmic interaction.

This model suggests that an extension of the paradigm modelled in this chapter that investigates human capacity to distinguish between live interaction and interaction with a recording may produce further insights into the dynamics of perceptual crossing and that, despite the minimalism of the simulated environment, simple action-perception strategies may bring about results similar to those reported from the double TV-monitor experiments (Trevarthen, 1979).

Indeed, De Jaegher, Di Paolo and Wood (personal communication) have recently replicated

and extended the experiments reported in (Auvray et al., 2008). Some preliminary results show that the hypothesis generated from Iizuka's extended model, i.e., that instability of one-sided interaction is not a consequence of the attachment of the lure in the original experiment. Furthermore, the results of the replication of the original study, apart from confirming the overall findings, appear to suggest that successful interaction relies on or triggers a coordination of strategies. Further analysis will be necessary to confirm these observations.

This case provides a nice example of a proper dialogue between empirical studies and simulation models that I outlined in section 3.6: An experimental study is modelled in a computer simulation, which increases our understanding of the data obtained, because the simulation produces results that go beyond our cognitive limits and prejudices and is, at the same time, easier to understand than the original phenomenon. From these results, an extended version of the experiments is generated, which is first investigated in simulation, leading to refined hypotheses and ideas. These ideas are then tested in empirical psychological experiments.

The experimental paradigm, despite its simplicity, is very rich and the possibilities for further research are open-ended and keep being explored experimentally. The simulation model presented in this chapter and (Di Paolo et al., 2008b) help to identify avenues for future research and to generate concepts and quantitative predictions to explain data gathered already or to be gathered in future experiments. The following chapter presents a simulation model of another experiment by the GSP that is a direct extension of the paradigm modelled in this chapter to a two-dimensional scenario. This study is the last model to be presented before the presentation of the interdisciplinary project on adaptation to sensory delays, for which I realised both the modelling and the experimental work with the objective that both would go hand in hand (chapter 8-12).

Chapter 7

ER in Dialogue with Perceptual Supplementation Research: Perceptual Crossing in a Two-Dimensional World

---

In the previous chapter, I presented the first application of ER modelling to Perceptual Supplementation (PS) research that I proposed in chapter 3. It has generated a number of interesting insights and hypotheses to be explored in data analysis and further empirical experiments. Despite the simplicity of the experimental paradigm, the investigation of perceptual crossing in a minimal virtual environment is intriguing and generates important fundamental insights and proofs of concept, both in their own right and in the light of the simulation modelling results. The GSP have conducted another experiment that is a direct extension of the original perceptual crossing study (Auvray et al., 2008) to the two-dimensional scenario and whose results have not been published yet (Lenay and Stewart, personal communication 2007/2008).

This chapter presents an Evolutionary Robotics model of this experiment that aims, in particular, at elucidating the role of human arm morphology in the generation of the quantitative properties of the recorded data. The results from this model have been published in (Rohde & Di Paolo, 2008). The work presented here chronologically followed the second part of this dissertation that combines experimental and modelling work to study adaptation to sensory delays (chapters 8-12). However, conceptually speaking, it is much more related to the previous chapter because both the experiment modelled and the model itself are a direct extension of those presented.

The following section 7.1 quickly introduces the purpose of the extended experiment and model, which are then described in section 7.2. I compared three different types of artificial agent on the task and found that the dynamical principles that govern the task are independent from agent bodies. The realisation of these principles is variable and depends on agent specific sensorimotor invariances. Such variability in evolved solutions includes the evolution of one-dimensional oscillation along a line in a simulated arm agent, a kind of behaviour that had been observed in the participants in the experimental study as well. The results are presented in section 7.3 and discussed in section 7.4.

7.1 Background: Perceptual Crossing in a Two-Dimensional Environment

Having investigated and analysed the dynamics and principles of perceptual crossing in a one-dimensional scenario (see chapter 6 and (Auvray et al., 2008)), Lenay et al. (personal communication) extended the experimental set-up to a two-dimensional virtual toroidal environment. With this modified set-up, the group wanted to test whether the experimental results transfer qualitatively or quantitatively from a one-dimensional to a two-dimensional scenario, which is by no means guaranteed: The sensorimotor contingencies afforded by the simulated toroidal environment are more complex and different from those in the one-dimensional set-up.

Some preliminary results from their study are that the results are indeed surprisingly similar to the results obtained in the one-dimensional experiment. Not only do the results transfer qualitatively (i.e., 65% correct clicks), but also the quantitative aspects of the behaviour are remarkably similar. In particular, interaction with an object or the other participant was realised by moving rhythmically back and forth along a line, reducing action to just one dimension, even if both dimensions were explored during search.

When learning about these results, my intuition was that the rhythmic one-dimensional interaction is related to the morphology of the human arm. The simulation model presented in this chapter, amongst other things, aims to establish the role of human arm morphology in the constitution of quantitative aspects of behaviour. Therefore, I modelled a simple simulated arm agent and compared it to two other kinds of artificial agents, i.e., a two-wheeled robotic agent and an agent that generates a velocity vector anchored in Euclidean space, similar to a joystick (called the 'Euclidean' agent; details of the environment, tasks and agents modelled in section 7.2). This latter type of agent can be seen as directly extending the agent architecture used in the model of the one-dimensional version of the experiment, whereas the sensorimotor couplings of the other two agents in the task are radically different.

The objective of comparing these different kinds of controllers is to identify common dynamical principles that derive from the task and the environment and that are relatively independent of embodiment and to distinguish them from qualitative and quantitative aspects of behaviour that are specific to a certain type of body or sensorimotor coupling.

The results point out some interesting common principles and quantitative differences. For instance, one-dimensional oscillation along a line evolved in the Euclidean and the simulated arm agents but not in the two-wheeled agents. A very efficient strategy evolved, which is counter-intuitive and contrasts with the strategies employed by the human participants: agents establish stable interaction with the fixed lure and avoid the other agent, which results from a change in the fitness function as compared to the model for the one-dimensional version (chapter 6). Both this surprising strategy and the finding that one-dimensional rhythmical interaction can result from arm-like agent morphology increase out understanding of the dynamics afforded by the task and generate hypotheses to be explored in the analysis of the original experimental data.

7.2 Model

As it was the case in the model of the one-dimensional version of the experiment, the simulation used for the evolution of artificial agents was, apart from parameter details, identical to the one

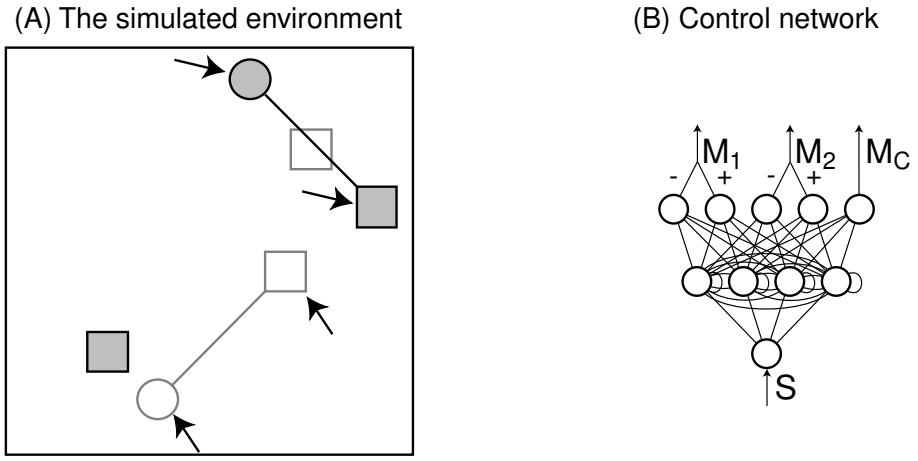

Figure 7.1: Schematic diagram of the simulation environment and control network. (A) The simulated environment with the two agents (circles), the attached lures (boxes attached with a line) and the fixed lures (boxes). (B) The control network.

used in the original experiment.

The simulated environment is a (200×200) virtual torus, i.e., a plane that wraps around in both dimensions. In this plane, there are six different objects. Two circular simulated agents of diameter 20, two mobile lures that are attached to the agents (at a fixed distance and angle) and two fixed lures that are statically installed at $(50,50)$ and $(150,150)$ respectively (see figure 7.1 (A): the agents are the circular objects, the attached and fixed lures are depicted as boxes in this and the other figures, even though they are also circular of diameter 20 in the simulation). The attached lures shadow the trajectories of each of the agents at a distance of 93 units, being attached in perpendicular directions.

The only sensory signal S that the agents receive is a touch signal, i.e., if the distance d between the agent and something else is $d < 20$, an input S_G (sensory gain, evolved) is fed into the control network. Each agent can only perceive the other and one of each kind of lure, i.e., the dark agent can perceive all light objects in figure 7.1 (A), but not the dark ones, and vice versa, in order to make it impossible that interaction between the agents is mediated by another object that both agents perceive at the same time.

In order to investigate the role of morphology in the strategies evolved, and in particular the role of arm morphology, three different types of agents were evolved (specification below). For purpose of comparison, all three kinds of agent are controlled by structurally identical CTRNN controllers (compare chapter 3, equation (3.2) with one input neuron, four fully connected interneurons and five output neurons (figure 7.1 (B)). Four of the output neurons regulate the two motor outputs $(M_1 = M_G(\sigma(a_{M1}) - \sigma(a_{M2})), M_2 = M_G(\sigma(a_{M3}) - \sigma(a_{M4})), M_{1,2} \in [-M_G, M_G]$ with M_G being the evolved motor gain. These outputs are interpreted as $v_{l,r}$, $v_{h,v}$ or $\omega_{e,s}$ for different agents respectively (see below). The task is to interact with something and correctly classify if the object encountered is either of the lures or the other agent. The fifth output neuron generates the classification signal M_C to indicate whether interaction is with another agent (output $M_C > 0.5$) or with one of the lures (output $M_C \leq 0.5$).

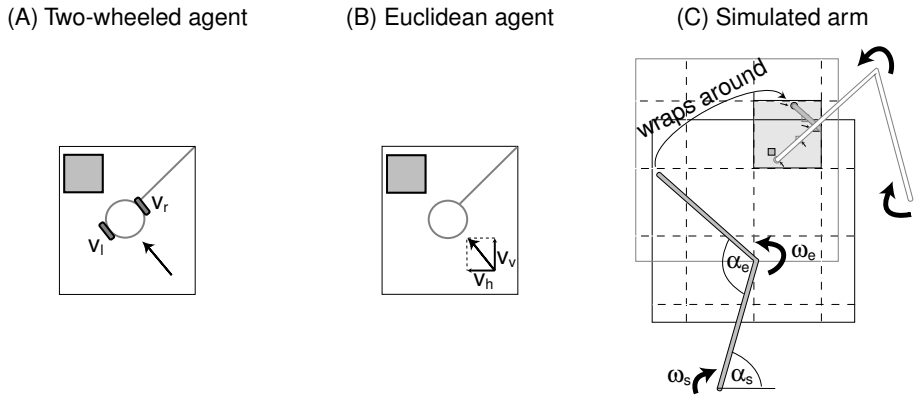

Figure 7.2: Schematic diagram of the different types of agents evolved. Diagrams of the two wheeled agent (A), the agent moving in Euclidean space (B) and two simulated arm agents, with the space in which they can act (C).

The three agent types evolved where

- *Two-wheeled agent.* The two-wheeled agent generates the velocity $v_{l,r} = 20M_{1,2}$ for each wheel (figure 7.1 (A); velocities are specified in units/s).
- *Euclidean agent.* The agent that I call the 'Euclidean' agent generates a horizontal and a vertical velocity vector $v_{h,v} = 30M_{1,2}$ that are summed up to define a vector in absolute space (figure 7.1 (B)). This agent can be seen as the two-dimensional analogy to the agent generating left and right movement modelled in the one-dimensional model in chapter 6.
- *Arm agent.* A simple simulated arm with two segments of length 400 units that is steered through angular velocity signals $\omega_{e,s} = 0.05M_{1,2}$ to the elbow and the shoulder joint (see figure 7.1, (C)). In order to approximate the dynamics of human mouse motion, the arm agent is restricted in its movements in two ways: through joint stops $\alpha_s \in [0.1\pi, 0.6\pi]$ and $\alpha_e \in [0.2\pi, \pi]$ and through the delimitation of movement to an area of 600×600 units that represents the 'desk' surface (i.e., the area within which a human participant would move the mouse), whose bottom left corner is fixed at $(-200, 200)$ taking the shoulder joint as the origin. The desk area is translated randomly with respect to both the desk area of the other agent and the simulated virtual environment to avoid that agents evolve to meet in the middle of the desk.

A problem with the simulated arm agent was that it has no way of telling where with respect to its anchoring in absolute space it is, because it has no proprioceptive sensors that represent its joint angles or any other form of telling where it is and whether it is still moving or has run up to a joint stop. This is the reason why the arm agents did not evolve to a high level of performance (see section 7.3. A modified version of the arm agent with three sensory neurons that received the joint positions as additional inputs ($S_{2,3} = S_G\theta_{e,s}$ was evolved for purposes of comparison. I decided, however, not to put much energy in trying to resolve these problems because I found many of the original questions already addressed by the original defect set-up. Controllers for all three kinds of agents were evolved without sensory delays and with a 100 ms sensory delay.

I used the GA and parameters specified in section 3.3 ($r = 0.6$) to evolve agents (74 parameters) that were matched against clones of themselves in the task. I ran 10 evolutionary runs over 1000

generations for each agent body with and without delay. Parameter ranges were: $S_G, M_G \in [1,50]$, $\tau_i \in [20,3000]$, $\theta_i \in [-3,3]$ and $w_{i,j} \in [-6,6]$.

Each trial lasts $T \in [6000,9000]$ ms. The starting positions were random for the wheeled and the Euclidean agent and random within the centre area for the arm agent. The starting angle for the wheeled agents is random. For the arm agent and the Euclidean agent, the relative orientation of the agents to each other is random $\in \{\frac{-\pi}{2}, 0, \frac{\pi}{2}, \pi\}$. The fitness $F(i)$ of an individual i in each trial is given by the following function

$$F(i) = \begin{cases} 1 & \text{if } (d_s \leq D) \wedge (d_o > D) \wedge (M_C > 0.5) \quad \text{(true positive)} \\ 1 & \text{if } (d_s > D) \wedge (d_o \leq D) \wedge (M_C \leq 0.5) \quad \text{(true negative)} \\ 0.25 & \text{if } (d_o < D) \wedge (d_s < D) \quad \text{(ambiguity)} \\ 0.1 & \text{if false classification and } S > 0 \quad \text{(touch)} \\ 0 & \text{else} \end{cases} \tag{7.1}$$

where $D = 30$, d_o the distance to the closest of the two lures and d_s the distance to the other agent. Agents are tested on eight trials and fitness is averaged. This fitness criterion is conceptually different from the fitness criterion used in the one-dimensional version of the simulation model. It resembles the task posed to the human participants more closely, as I do not evolve to choose interaction with the other agent but instead to classify if an entity encountered is the other agent or not. Interestingly, this relaxation of the constraint to seek interaction with the other agents led to the evolution of a preference for interaction with the fixed lure, as discussed later on in this chapter.

7.3 Results

7.3.1 Evolvability

The wheeled agent and the Euclidean agent evolve to a much higher level of performance (see figure 7.3, (A)), with the best individual from the best evolutionary run achieving nearly perfect performance, whereas even the best arm agent clearly stays below a fitness of 50% (figure 7.3, (B)). Part of the reason for this discrepancy is that the arm agent does not have means to orient itself in space. For the Euclidean and the wheeled agents, there are simple strategies (fixed motor outputs) that allow them to scan the space (i.e., to go into a non-horizontal or non-vertical direction for the Euclidean agent or to go around in circles/spirals/curves for the wheeled agent). The arm, however, will run up to a joint stop or the edge of the desk surface if it applies any constant angular velocity to any of the joints without receiving any sensory feedback about whether it is still moving or not. This disadvantage made evolution of the arm much more difficult and subject to randomness than those of the wheeled or Euclidean agent (cf. figure 7.4, bottom left).

I evolved agents with proprioceptive inputs (joint angles) for comparison and they immediately achieved much higher levels of fitness (population average/best after 1000 generations in 10 runs: 0.33/0.70) and evolution was less noisy (figure 7.4, bottom right). Despite this patch of the model, the arm agent did not evolve to near perfect fitness like the wheeled agent and the Euclidean agent did. However, the question I wanted to address with the model, i.e., the role of arm morphology in the constitution of rhythmical one-dimensional trajectories, could be addressed using the simulation results obtained even in the non-proprioceptive model, even though further experiments to

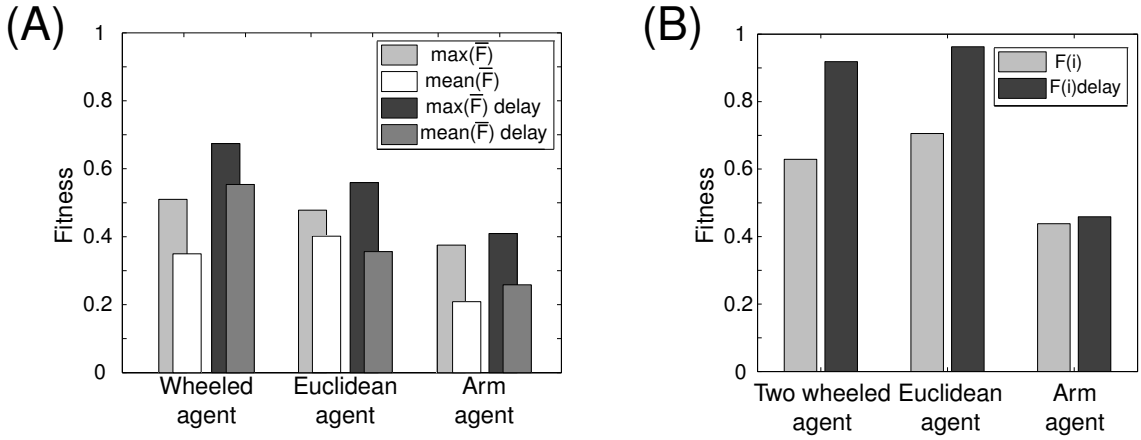

Figure 7.3: (A) Population fitness average $\bar{F}$. Mean and maximum from 10 evolutionary runs, with and without delay. (B) Performance average across 100 evaluations for the best individual from the best evolution. Dark: 100 ms delay, light: no delay.

improve the arm model are clearly an interesting avenue for further experiments.

All agents evolved to a higher level of performance with delays than without (see figure 7.3, (A)), as already observed for the one-dimensional scenario presented in the previous chapter. Figure 7.4 (top) depicts typical fitness evolution profiles for the wheeled agents without (left) and with (right) sensory delays. This shows that evolution without delays quickly converges to a non-optimal solution (local maximum), whereas evolution with delays converges as quickly to a near-perfect solution. The nature of this evolvability benefit provided by sensory delays is discussed in more detail in the following section 7.3.2 and relates to the evolution of rhythmic interaction behaviour as opposed to search-and-stop behaviour.

7.3.2 Behavioural Strategies Evolved

Irrespective of agent body, two large classes of behaviour dominate the fitness landscape for the perceptual crossing task. The more successful strategy (1) is to avoid any mobile objects, search for the fixed lure, interact with it and always output 'no' ($M_C \leq 0.5$). This strategy leads to up to perfect fitness. Even though perfectly viable, this strategy is rather un-intuitive (tongue-in-cheek, I termed this strategy 'autistic' in (Rohde & Di Paolo, 2008)). It also clearly contrasts with the participants' behaviour, who avoid the fixed lure and seek interaction with each other. The second predominating strategy (2) is to interact indiscriminately with any entity encountered and to output 'yes' ($M_C > 0.5$) constantly. This strategy yields a fitness of up to ca. 40%. It appears, thus, that what evolved were preferences rather than discriminatory capacity; even if agents evolved to interact with all kinds of objects (strategy (2)), it appears to be more advantageous to exploit the slight combinatorial advantage of a permanent 'yes' answer over a permanent 'no' answer and not to intend a discrimination based on sensorimotor interaction with an object. The arm agents nearly exclusively evolve strategy (2), whilst the Euclidean and the wheeled agent evolve strategy (1), frequently passing during evolution through a phase of strategy (1). Only four agents (one arm, one wheeled, two Euclidean) evolved a contingent classification output triggered by stimulation (e.g., say 'yes' if you touch something, in case you run into the other last minute and 'no' if

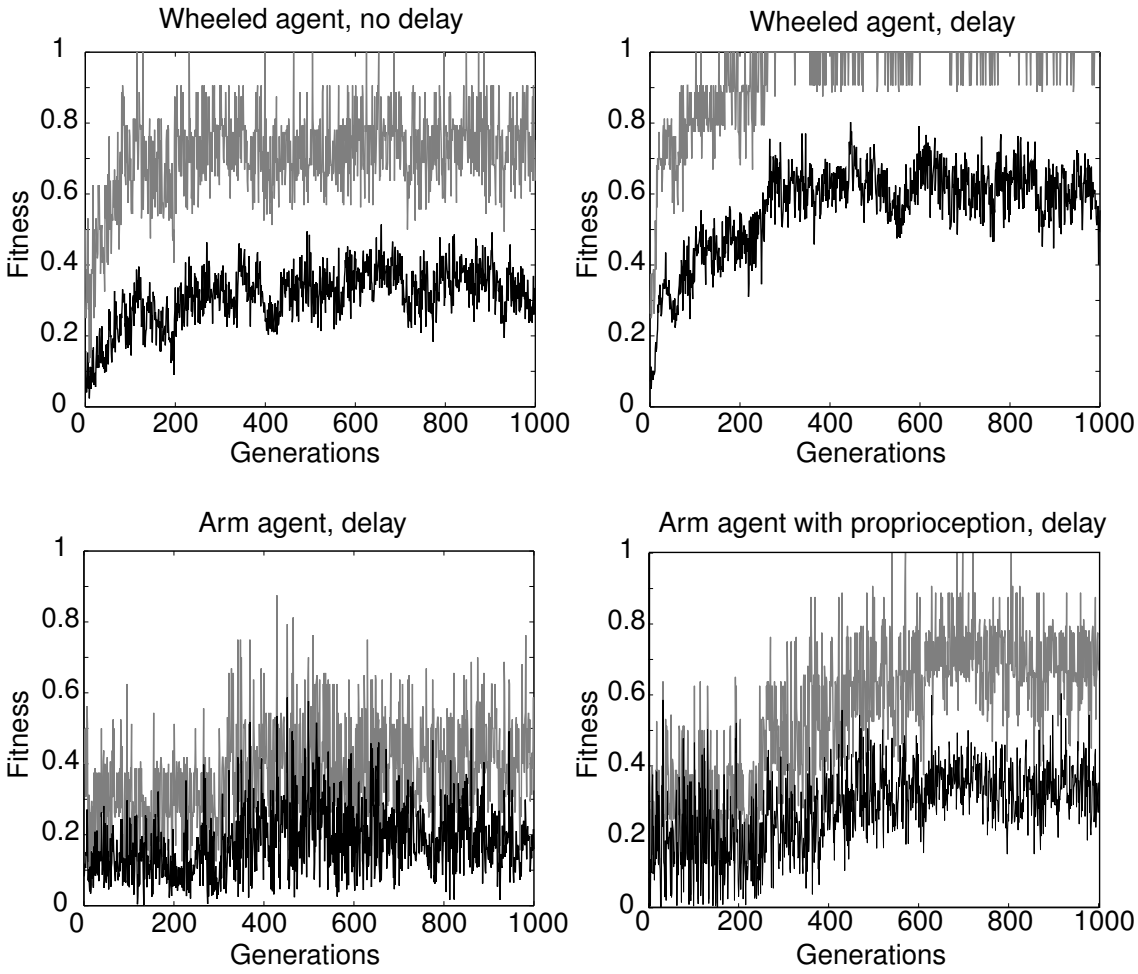

Figure 7.4: Example evolution profiles for different agents and parameters, black: population average, grey: population best. Top left: wheeled agent, no delay (search-and-stop solution. Top right: wheeled agent, delay (rhythmic solution). Bottom left: arm agent delay (noisy). Bottom right: arm agent with delay and proprioception (less noisy).

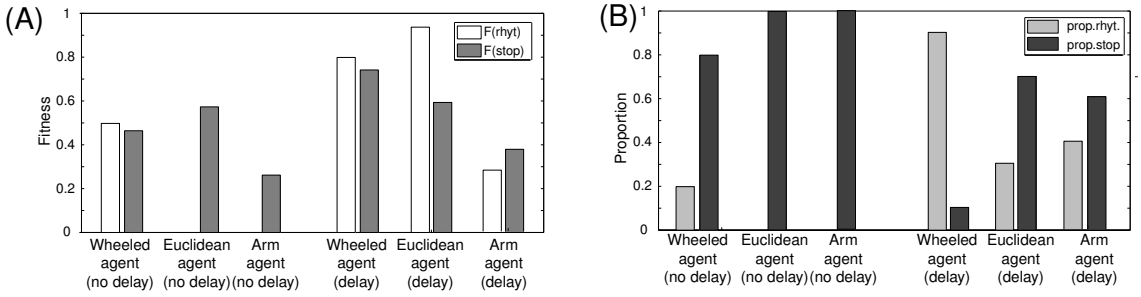

Figure 7.5: Average of populations in which rhythmic behaviour was evolved and correlated fitness. (A) Fitness for rhythmic solutions (white) is on average much higher than that for non-rhythmic solutions (grey). (No rhythmic action was evolved for Euclidean or arm agents without delay; note that the measure for rhythmicity is an approximation as explained in section 7.3.2.)

stimulation continues over an extended period of time) additionally to a behavioural preference. This preference for interaction with the fixed lure contrasts with the experimental results and also with the synthetic results from the model presented in chapter 6, in which preference for live interaction and had been presupposed and built into the fitness function.

Both strategy (1) and strategy (2) involve localising another entity and staying close to it. Staying close can be realised, in principle, by rhythmical interaction with the target or by simply stopping where the stimulation does not cease. It appears that rhythmic behaviour is more adaptive: if we define, as an approximation, rhythmic behaviour as activity confined to a radius of $d = 50$ around an entity during the last second of a trial with at least five inversions of sensory state, we find that within each agent type for which both oscillating and non-oscillating solutions evolved, the oscillating ones were on average 9% more successful (see figure 7.5 (A); note that, due to the noisiness of arm evaluation, some of the rhythmic solutions evolved in arm agents with delay were not recognised by this approximate measure).

The reason for the adaptive advantage of rhythmic strategies is that an agent evolved to simply stop is clueless where the stimulant has disappeared to if stimulation suddenly ceases. Such unexpected cessation can happen, e.g., when crossing an object at an unfortunate angle. It will start the search for sensation anew. An agent that interacts with an object rhythmically is moving repeatedly towards and away from its boundary and therefore has at least some capacity to relate its actions to the sensation of the object, inverting the effect of an action that makes stimulation go away. Thereby it establishes how it spatially relates to the object. With this minimal spatial interaction, if stimulation unexpectedly disappears, the agent has at least the possibility to go into the direction of the last stimulation, which increases the probability to re-encounter the lost object.

As in the one-dimensional version of the model, integrated sensory stimulation over time that represents perceived size of the object is crucial for distinguishing fixed or mobile objects. In order to test this hypothesis, I varied the size of the objects in the virtual environment (as in the one-dimensional version of the model). If the size of the other agent is doubled or the size of the fixed lure is divided by two, the fitness of the arm agents, who do not make the distinction between mobile or fixed objects drops only marginally altered 0.33 to 0.5/0.28 for doubled/halved respectively. These differences can be explained solely through the increased or decreased probability

of making contact with another entity in the first place. For the Euclidean and wheeled agents that seek interaction with the fixed lure only, fitness deteriorates completely with these alterations, dropping from 0.69 to 0.11/0.07 and from 0.79 to 0.08/0.07 respectively, showing that their discriminative capacity is severely impaired by the alteration of size and the subsequent differences in integrated duration of stimulation during interaction.

Sensory delays seem to be crucially involved in bootstrapping the evolution of this kind of solution: rhythmic behaviour as defined above evolved to occur at least once in 10 trials in 2 of the 30 best individuals evolved without delays and in 16 out of 30 best evolved individuals with delay. With a delay, objects are only registered once an agent (in all three conditions) already shot past it. This forces agents to stop and return to the locus of stimulation, which is a more advanced behaviour and helps to overcome a local maximum in the fitness landscape, i.e., to stop upon any stimulation and start the search anew if stimulation unexpectedly ceases, which again bootstraps the evolution of effective and active perceptual strategies (cf. figure 7.5 (B)).

The exact realisation and behavioural dynamics vary quite significantly between conditions, as analysed in the following subsections for the agents evolved with delays. My objective with this model was to explore the space of possible solutions and a detailed investigation of example agents (best agents evolved with delays) will help to understand and clarify those. In particular, it has been observed that, across agent bodies, two behavioural phases, search phase and interaction phase, can be realised variably and independent of each other.

7.3.3 Two-Wheeled Agent

Wheeled agents evolved a variety of strategies to search for objects in the toroidal environment: some shoot off in one direction, others drive around in large circles, arches or spirals. When an object is encountered, interaction is either initiated immediately, or, alternatively, the agent backs off and comes back to see if the stimulating object is still there, a strategy which contributes to localising the fixed lure rather than the other agent or the attached lure in the 'autistic' solution to the task.

All wheeled agents evolved to drive in circles (of variable size) around the encountered entity, most of them aiming at a distance from the object that makes stimulation rhythmically appear and disappear. Figure 7.6 depicts a sample behaviour of the best agent evolved with average fitness $F(i) = 0.92$. Agent 1 (black solid line) is in stable interaction with the fixed lure throughout the time period depicted. Agent 2 (dotted solid line), on the other hand, is momentarily trapped in an interaction with agent 1's attached lure (black dotted line and grey solid line, t = [500,1500]). The interaction does not stabilise, because stimulation through the mobile attached lure is too intermittent, even though it is maintained over a number of crossings. The agent thus eventually abandons the lure, passes the other agent twice (both times touching it very shortly and, consequently, not performing a complete return trajectory, and then finds the fixed lure. This strategy only fails in very exceptional cases in which interaction with a mobile entity is phase-locked in a way that resembles interaction with a fixed lure.

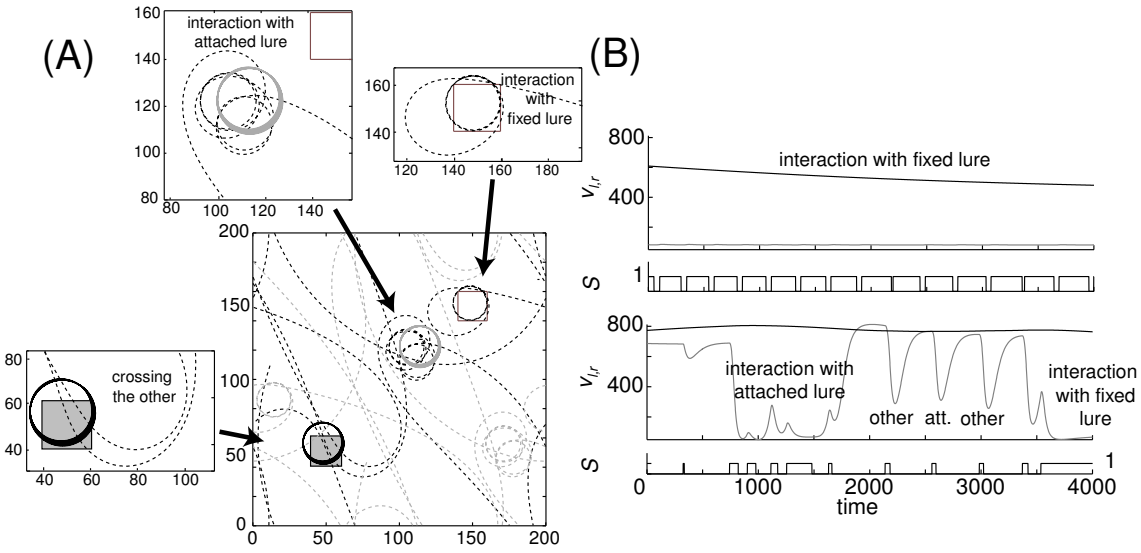

Figure 7.6: Example trajectory and sensorimotor diagram for the best wheeled agent evolved. (A) The trajectory over the entire time period (large square) and local trajectories during significant sub-behaviours enlarged (small squares). Agent 1 solid line, agent 2 dotted line; agent movement black, movement of attached lure grey. (B) Sensorimotor diagram $v_{r,l}$ and S (rectangular) during the behaviour depicted in (A). Agent 1 top, agent 2 bottom.

7.3.4 'Euclidean' Agent

An architectural advantage that the Euclidean agents have is that the direction of their movement is anchored in Euclidean space. This inbuilt 'sense of direction' allows them to scan the space by applying a constant motor output, producing straight lines on the torus that wrap around it in a tight spiral (see slightly displaced lines in figure 7.7 (A); best Euclidean agent evolved with average fitness $F(i) = 0.96$). This is an extraordinarily efficient search strategy. Only two agents evolved to start search in a large curve.

Figure 7.7 depicts the behaviour of the best agent evolved: if either of the agent encounters a mobile entity that moves perpendicularly, the stimulation is so short that the velocity is only minimally decreased (kinks in trajectories) and not even repeated crossing is initiated. The Euclidean agents exploit their absolute sense of direction because it constrains the angles at which they could possibly meet, due to the limited number of relative starting orientations. Agents move either exactly parallelly (unlikely to meet) or exactly orthogonal (very short stimulation).

Once contact with the fixed object is made, half of the agents evolve to simply stop upon stimulation, rather than to engage in rhythmic interaction. This tendency probably accounts for the slight population disadvantage of the Euclidean agents as compared to the wheeled agents. The other half evolve to rhythmically interact with the fixed lure along one dimension, implementing the 'autistic' strategy (1) to the task by making stimulation continually appear and disappear.

A behavioural pattern that only evolved in some of the Euclidean agents is to systematically destabilise even interaction with the fixed object, by slowly grinding past it (for stop solutions), or by moving further away with each oscillation (for rhythmic solutions). This strategy makes it possible to avoid interaction with mobile objects more efficiently and also breaks interaction

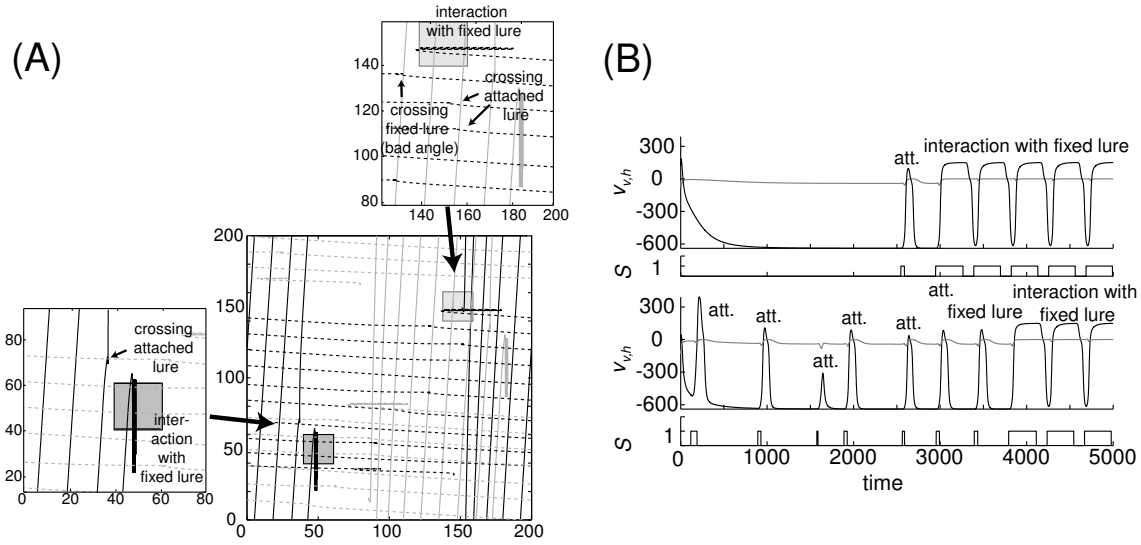

Figure 7.7: Example trajectory and sensorimotor diagram for the best Euclidean agent evolved. (A) The trajectory over the entire time period (large square) and local trajectories during significant sub-behaviours enlarged (small squares). Agent 1 solid line, agent 2 dotted line; agent movement black, movement of attached lure grey. (B) Sensorimotor diagram $v_{h,v}$ and S (rectangular) during the behaviour depicted in (A). Agent 1 top, agent 2 bottom.

in the rare occasions where interaction with a mobile object resembles interaction with the fixed lure. Even if this technique leads to the occasional loss of the fixed lure, due to the very efficient search strategy of the Euclidean agents, the probability to find it again quickly is very high. This strategy, as the strategy employed by the successful wheeled agents, is very effective and fails only in exceptional cases.

7.3.5 Arm Agent

As mentioned earlier, the arm agents evolved to much lower levels of fitness. This disadvantage is probably largely due to the fact that, other than the other two types of agents, arm agents do not have an easy way of exploring the environment. Without proprioceptive feedback, the agent has no way of telling where it is and whether it is still moving or has run up to a joint-stop or the edge of the desk. No constant output will yield any efficient search behaviour.

The agents evolved to either approach the desk edge in a large arch and then grind down the edge or to quickly go to one extreme arm position (neuron with fast τ) and then scan back in a large curve (neuron with slow τ). Both these scan behaviours fail if no object is encountered the first time this movement is executed. This enters randomness into the fitness evaluation, as behavioural success largely depends on appropriate objects lying on the path of the reflex-like movement executed by the arm. This makes evolution very noisy, as mentioned in section 7.3.1.

From the original series, only one agent evolved a scanning behaviour that goes beyond the execution of one blind swaying movement: it makes use of a neural oscillator as central pattern generator (CPG). The trajectories it generates and the sensations and motions over time are depicted

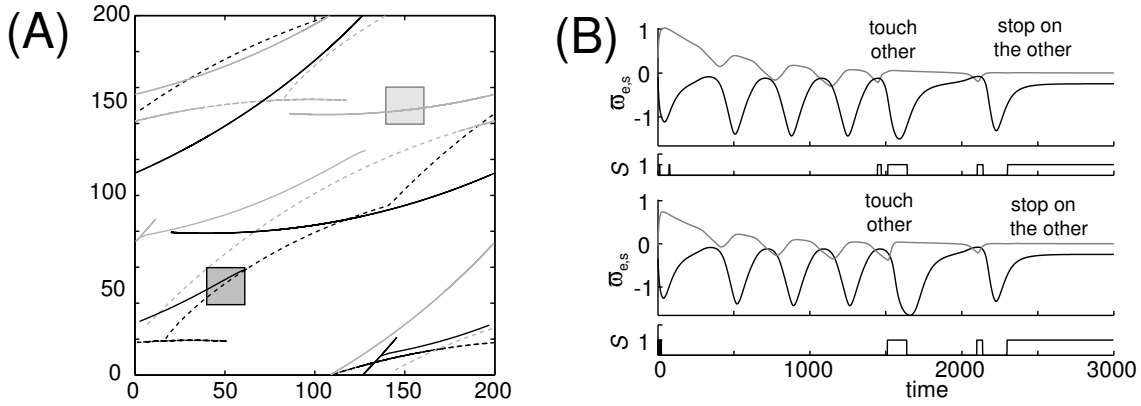

Figure 7.8: Example trajectory and sensorimotor diagram for an arm agent that evolved a neural oscillator as central pattern generator. (A) The trajectory over the entire time period (large square). Agent 1 solid line, agent 2 dotted line; agent movement black, movement of attached lure grey. (B) Sensorimotor diagram $\omega_{e,s}$ and S (rectangular) during the behaviour depicted in (A) Clearly shows the oscillatory outputs in the absence of sensory inputs. Agent 1 top, agent 2 bottom.

in figure 7.8.[1] This agent is the second best agent evolved, even though it has no sophisticated interaction strategy (i.e., sensation initiates the decrease of motor outputs to 0).

Nearly all arm agents evolve to rhythmically interact with any entity encountered (even if that is not always recognised by the criterion specified in section 7.3.2), making the sensory stimulation constantly appear and disappear. The best agent evolved with average fitness $F(i) = 0.46$ (see trajectory and sensorimotor diagram in figure 7.9) implements this kind of behaviour. Again, the rhythmic powering of one joint only leads to the *exact* inversion of the path just made.

As expected, the rhythmic activity in the arm agent leads to the production of near-straight oscillatory trajectories, as they were observed in human participants. The interesting aspect about this result is that, even though such trajectories did not evolve in all agent types (wheeled agents evolved to drive around in circles), it seems to be the arm-specific implementation of a general principle, i.e., the reduction of motion to oscillatory behaviour in one dimension of the output space only.

Looking at the behaviour and performance levels attained in the complementary evolution of arm agents with proprioceptive feedback reveals that, even though solutions do have higher fitness on average, arm agents with proprioception evolve still strategy (2), i.e., indiscriminate interaction. The resulting interaction behaviour is, in many ways, similar to the behaviour evolved in successful arm agents without proprioception (figure 7.10 (A) and (B)), even if the localisation behaviour is more successful. The additional proprioceptive input mitigates some of the problems with noisy evolution and behavioural randomness associated with the impossibility of spatial orientation. It does, however, not lead to the evolution of perfect or near perfect solutions, such as strategy (1).

There are possibilities for further analysis of why this is so, and more ways of trying to further improve the arm agents' performance (such as longer evolution due to the larger parameter space).

[1] The trajectories generated are quite confusing, because during each oscillation, a part of the previous path is exactly inverted by inverting velocity on one joint and decreasing angular velocity on the other joint to 0. This visualisation problem is quite common for solutions evolved in arm agents and also characterises the solution depicted in figure 7.9.

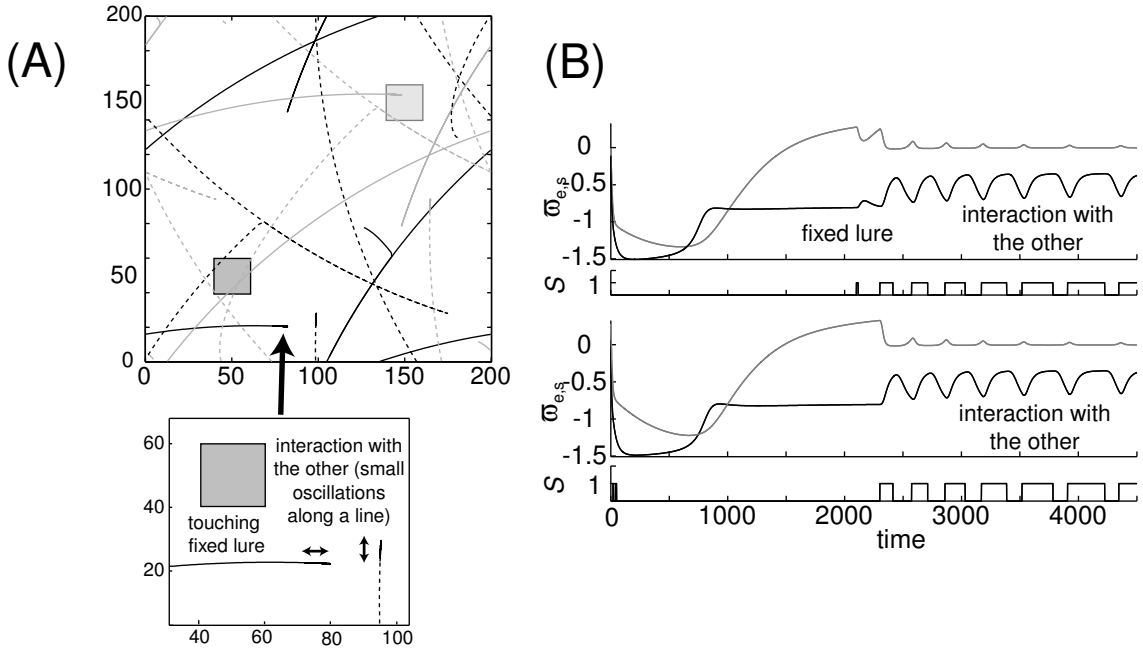

Figure 7.9: Example trajectory and sensorimotor diagram for the best arm agent evolved. (A) The trajectory over the entire time period (large square) and the trajectory during interaction enlarged (small square). Agent 1 solid line, agent 2 dotted line; agent movement black, movement of attached lure grey. (B) Sensorimotor diagram $\omega_{e,s}$ and S (rectangular) during the behaviour depicted in (A). Agent 1 top, agent 2 bottom.

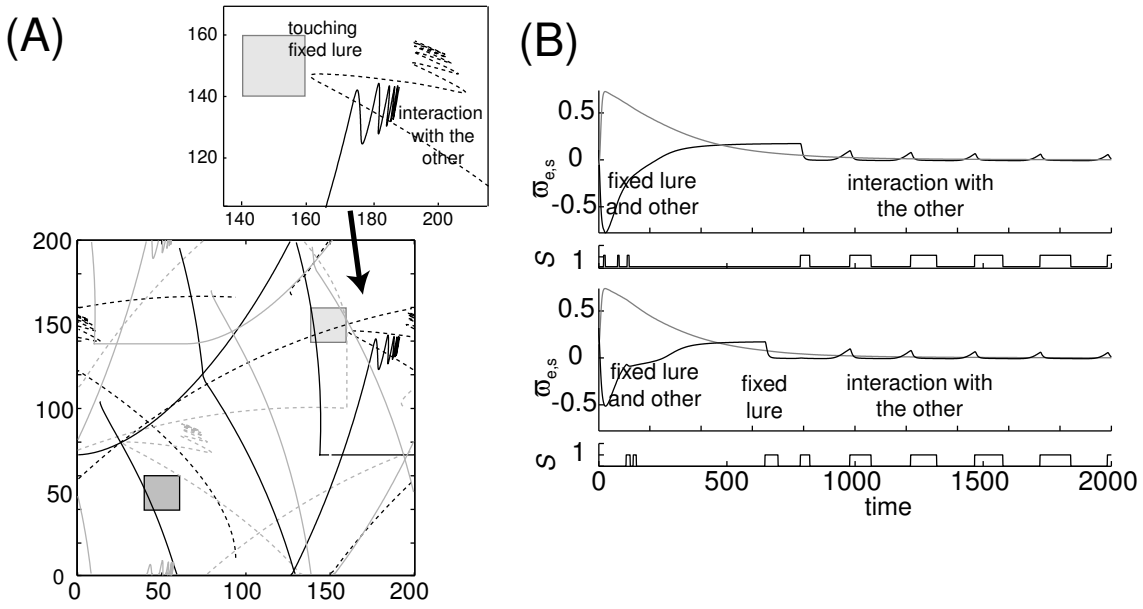

Figure 7.10: Example trajectory and sensorimotor diagram for an arm agent evolved with proprioceptive feedback. (A) The trajectory over the entire time period (large square) and the trajectory during interaction enlarged (small square). Agent 1 solid line, agent 2 dotted line; agent movement black, movement of attached lure grey. (B) Sensorimotor diagram $\omega_{e,s}$ and S (rectangular) during the behaviour depicted in (A). Agent 1 top, agent 2 bottom.

One of the main questions behind this model can, however, already be addressed with the sub-optimal results obtained. The results show how arm morphology produces oscillation along one dimension as the implementation of a general dynamical principle, i.e., rhythmic interaction along one dimension of motor space (see following discussion).

7.4 Discussion

A main result from this simulation model is that several dynamical principles govern the evolution of solutions to the modelled task. These hold across different agent bodies.

- The search space of possible strategies is dominated by two principal solutions: (1) Avoid mobile objects, seek interaction with the fixed lure and output 'no'. (2) Interact indiscriminately and output 'yes'.
- Strategy (1) is the more successful strategy and yields nearly perfect fitness.
- Solutions that rely on rhythmic interaction are on average more robust to perturbations because they facilitate spatial localisation of the stimulant and thus yield higher fitness than solutions that rely on stopping on top of a stimulant.
- During this rhythmic interaction, one motor signal implements the oscillation, the other one is frozen and serves to adjust behaviour if necessary.
- Evolution of the superior rhythmic solutions is facilitated by the introduction of a 50ms sensory delay.
- Two different behavioural modes that can be realised variably and independently are identified: search and interaction.
- Despite the quantitative differences in how the behaviour manifests in space and time, the sensorimotor diagrams displaying sensorimotor activation over time are of remarkably similar appearance.

Apart from these commonalities, there are different quantitative properties associated with the realisation of these dynamical principles across the different agent bodies.

- The realisation of search and interaction behaviour is strongly influenced by agent morphology and the sensorimotor couplings that characterise and constrain the space of possible solutions.
- In particular, search behaviour can be particularly efficiently implemented in the Euclidean agent and is extremely difficult to evolve in the simulated arm agent. The difficulty of evolving search behaviour implies a drastic disadvantage in overall evolvability for the simulated arm agents.
- Rhythmic interaction behaviour is realised differently in all three agent types. In particular, wheeled agents circle around the object encountered, whereas the arm agent and the Euclidean agent engage in one-dimensional rhythmic interaction. In the Euclidean agent this implies oscillation along either the absolute vertical or horizontal dimension, while in the arm agent, oscillation of either of the joints results in slightly curved oscillations along the orientation of the arm.

These simulation results support my expectation that arm morphology plays a role in the one-dimensional rhythmic interaction observed in human participants, as the arm-specific implementation of a more general dynamical principle governing the task. They predict that in the gathered data, observed oscillations should be orthogonal to the orientation of the arm and that this oscillation should serve to establish rhythmic interaction with the encountered object or participant.

An interesting parallel with the one-dimensional version of the simulation study is that, again, sensory delays improve evolvability because they bootstrap the evolution of oscillatory scanning behaviour possible. This result suggests an investigation of dependencies between sensorimotor latencies and frequency of oscillation in the experimental data, just like the results presented in chapter 6. Also, integrated sensory stimulation time and how it correlates to perceived size of the object/agent appears to play a key role in distinguishing the fixed lure from the other agent. As in the one-dimensional version of the experiment, this synthetic result predicts that integrated stimulation time correlates to the decision made.

A difference between the experimental result and the modelling results presented in this chapter is that experimental participants seek interaction with the other participant, whereas, in the simulation the dominating strategy (1) is an 'autistic' strategy in which agents avoid each other and seek for the fixed lure. This surprising result also contrasts with my earlier simulation model, for which I had required agents to seek interaction with one another, presuming a preference for live interaction. From these results, I had concluded that perceptual crossing is, given the task, a nearly inevitable result from the mutual search of the agents/participants for each other (see chapter 6), even if this simulation already hinted towards the difficulty to avoid the static lure. In the light of the present simulation results it becomes clear that, leaving aside motivational factors (such as boredom), the dynamics of the task do not favour perceptual crossing, but much rather interaction with the static lure, and that perceptual crossing is established despite this strong basin of attraction.

Feeding the results back to the researchers of the GSP who have conducted the experiment (they are still analysing their unpublished results), my simulation efforts have again been appreciated. In particular, they found that the simulation results clarified the role of morphology in the recorded behaviour and the evolution of autistic behaviour pointed them to an implicit presupposition in their formulation of the task. Further simulations to investigate the dynamical principles of the task have been suggested (Lenay, personal communication, Feb. 2008).

The four simulation models presented in the previous chapters have addressed different kinds of research questions. The model of linear synergies (chapter 4) aimed at exploring a concept from human motor control research in strongly minimised and idealised settings, in order to generate hypotheses for further experiments and to generate proof that the postulated principles can work in theory. In a more philosophical endeavour, the model of value system architectures presented in chapter 5 caricatured a neural architecture proposed as a mechanism for general behavioural adaptation, pointing out implicit premises underlying the proposed principles. The previous chapter and this chapter have applied minimal ER modelling to findings from PS research, as I proposed in chapter 3. As argued in section 3.6, the close match between experiment and simulation allows a much stronger analogy between model and experiment that serves to generate quantitative predictions about experimental data from previous and future experiments, alongside with the more

abstract proofs of concept and counter-intuitive insights resulting from ER as a tool for thinking in theory-building.

All four models have generated valuable contributions to the problem area they address. Furthermore, the models presented in chapters 4, 6 and this chapter have also been value-*ed* by the experimental researchers working in the field. These findings alone show already that ER modelling is a fruitful method for the investigation of human level behaviour and cognition. Furthermore, concerning my personal research interests, while the model of motor synergies addressed a question that was more 'low level' than the questions that most interest me, and the model of value systems was more abstract and remote from scientific practice than I would like to work, the models on perceptual crossing in one and two dimensions directly address intriguing questions of high level human cognition and perceptual experience. These results give reason to be optimistic about the possibilities of joint PS experimental work and ER simulation modelling work as mutual sources of inspiration.

The following chapters (8-12) present the results from the interdisciplinary study of adaptation to sensory delays and perceived simultaneity, which I refer to as the second part of my dissertation. It puts to work the framework I propose in section 3.6 to combine experimental, synthetic and experiential methods. Chapter 13 assesses the methodological significance of all the models presented in the scope of this dissertation as concerns their contribution to the study of human behaviour and cognition.

Chapter 8

The Sensorimotor Basis of Temporal Experience

This chapter marks the beginning of the second part of my dissertation, i.e., it is the first of five chapters on the interdisciplinary investigation of adaptation to sensory delays and perceived simultaneity. As such, it serves as a 'second introduction' chapter, it is entirely conceptual - a super-sized version of the introductory sections in the previous results chapters 4 - 7. I present some general thoughts and results on the sensorimotor basis of time cognition, before I review the relevant research context in order to frame the question addressed with the interdisciplinary project. Chapters 9-11 present the experimental and simulation results that are evaluated with respect to the question here framed in the discussion chapter 12. The conclusion chapter 13 evaluates the project on adaptation to sensory delays along with the previous simulation models and results presented with respect to the methodological theme of this dissertation framed in chapters 2 and 3.

The perception of time is a curious problem and possibly one of the hardest in the study of human cognition. Findings in embodied time perception from disciplines as diverse as phenomenology, neuroscience, anthropology, psychophysics, philosophy, linguistics and psychology are presented. Each of the sections below would deserve an entire book; I raise more questions in this chapter than I provide answers. Despite the broad range of authors discussed in this chapter, I am more than aware that important if not crucial perspectives are left out or misrepresented, and that my conclusions are likely to be either naïvely wrong or stating what others before me have found out more quickly and described in better words. I believe this is an inevitable problem when dealing with a question like time perception, which is a phenomenon nearly as big and interdisciplinarily studied as mind itself. In my defence, I have to say that many important contributions (such as Kant's (1974, recent edition; German original published in 1787) and Heidegger's (1963, recent edition; original published in 1927)) are not very accessibly written and probably unknown to a large proportion of contemporary cognitive scientists. In order to grasp the gist of my experiments and my attempt to explain at least one aspect of how our experience of time is constructed, I think it is important that I describe my various sources of inspiration and how I interpret them, how they shaped my view on time itself and how this conception is reflected in the approach I take to the problem studied.

I start gently by decomposing Cartesian intuitions about what the experience of time is and how

this view relates to traditional approaches in Cognitive Science to explain time perception (section 8.1). For the largest part of this chapter (section 8.2), I present the work of other thinkers and scientists in order to oppose the traditional and naïve view with the multitiered and rich reality of time perception which forms the starting point for an integrative enactive study of time perception. In its most basic form, a temporal dimension is crucially and irreducibly attached to the flow of consciousness as such, as many authors since Kant (1974, original in 1787) and possibly before have realised. At the same time, and probably as a consequence of the ubiquity and importance of temporal experience, time as such is one of the most abstract and cognitively high level constructs that the human mind reliably develops. What is it in our bodies and our interactions with the world that gives rise to the peculiar categorisation of encounters into those that are present, those that are past and those that are future? How do we come to impose an absolute and irreversible order relation on all the events of our world? How do we distinguish events (i.e., temporal entities) from objects (i.e., spatial entities)? Finally, in section 8.3, I home in on the problem of adaptation to sensory delays and simultaneity perception, introducing previous work in the area, my hypothesis and the research question addressed in the interdisciplinary study. This last section can be seen as a proper introduction to the experimental and simulation study presented in the following chapters.

8.1 Newton Meets Descartes: The Classical Approach

What is a caricatured naïve stance towards time cognition? Crudely speaking, it assumes that there is an objective time in the world, a Newtonian time arrow, that imposes a global order on events (before, at the same time, after) and defines absolute temporal distances between temporal events (a day before, five seconds after). A naïve representationalist and objectivist perspective on time cognition assumes then that time cognition is basically about trying to properly represent this real physical time arrow in our mental recreation of the world.

This approach has indeed been implemented in early AI systems and models of time cognition. They use, e.g., temporal logic that extends propositional logic to include time or tense stamps for each proposition (Allen, 1984). Similarly, indexicality with time stamps is used in formal semantics to disambiguate temporal language (Heim & Kratzer, 1998). The advantages of this view are a) that it appeals to our intuition of what time is and how it works and b) its simplicity. The disadvantage, however, is that with this view, one runs into three entire classes of drastic problems that I describe in the following in a little bit more detail: ontological problems about the nature of real time; technical problems about computers acting in real time; a failure to account for the phenomenon of mental time.

The objectivist premise that the absolute flow of time is real and basically Newtonian has been most severely injured from within the discipline of physics: first by Einstein's theory of relativity, then by the theory of quantum physics. In one of Einstein's straight forward thought experiments, he argues that the time accelerates or slows down if seen from different inertial systems that are at very different velocity relative to each other. A from the viewpoint of Newtonian physics paradoxical relativity of the order of events can arise from this difference in inertial system, given that the velocity of light is constant. Therefore, no absolute coordinate system or measurement of absolute time (which always involves a physically extended process) can exist. This entrance of the observer-scientist, the measurer, into the story of the very nature of time and the universe, was

pushed even further in quantum physics through the discovery of Heisenberg's uncertainty principle, which, in a crude simplification, says that the act of observation/measurement can change the outcome of an experiment. As such, these theories question only the Newtonian world view, not the general objectivity of time or physics, which leads to strange epistemological dilemmas and paradoxes (e.g., Schrödinger's cat). Bitbol (2001) argues how the tensions between quantum mechanics and dualist theories of knowledge can be resolved from a constructivist perspective. In a more objectivist language, the upshot has also been summarised by the early Wittgenstein: "We cannot compare a process with the 'passing of time' - which does not exist - but only with another process (for instance, the motion of a chronometer). Therefore, the description of the passing of time is only possible through reference to another process" (in Macho, 1996, p. 161 recent edition; TLP originally published in 1922 (TLP 6.3611)).[1]

This is, obviously, not to say that the construction and usage of clocks or the metaphor of an absolute time arrow are not useful. Indeed, dynamical systems theory, which is, as I argue in chapter 3, one of the prime mathematical and scientific tools for the enactive approach, usually employs only Newtonian time as a variable. It just has to make explicit that this useful construct does not possess any kind of ontological priority or reality over the rest of our useful mental constructs and, therefore, it requires explanation like all the others.

The second point is about the problems that GOFAI systems have with acting in real-time. These have been described by critics of the computationalist paradigm many times and have already been addressed in chapter 2. A system that exists in time and aims to represent the passing of time gets into trouble coordinating the internal and external time arrow. As Cantwell-Smith (1996) points out, in the case of a clock, this coordination is all it does and the closer the clock comes to mimicking the natural processes that were chosen to define temporal units, the better the clock. In the case of a digital computer, things are more difficult, because the formal language in which it is defined (automata theory) disregards real time, which means that any Turing machine can be instantiated in different ways that are temporally contingent. The implicit premise in Turing's (1950) 'Computing Machinery and Intelligence' is that exact timing is irrelevant to intelligence. This premise has been pointed out and refuted many times by different authors (for instance: Cantwell-Smith's criticism that "[traditional models of inference] take the temporality of inference to be independent of the temporality of the semantic domain" and that these need to be at least partially coordinated (Cantwell-Smith, 1996, p. 259); van Gelder's diagnosis that the computational hypothesis treats "time as discrete order" rather than a real-valued variable in his plea for the dynamical hypothesis in Cognitive Science (van Gelder, 1998, p. 6); Harvey et al.'s observation that computational systems are "a rather specialised and bizarre subset" of dynamical systems which are characterised by the fact that "updates are done discretely in sequence, with no direct references to any time interval" and are thus instantiated with accidental real-time properties (Harvey et al., 2005, p. 6)). All these authors come to the same conclusion: the need for embodiment and embeddedness and a formalism that unifies model-external and model-internal time (which is naturally granted in DST[2]). This realisation is already half the way towards an

[1] My translation: "Wir können keinen Vorgang mit dem 'Ablauf der Zeit' vergleichen - diesen gibt es nicht - sondern nur mit einem anderen Vorgang (etwa mit dem Gang des Chronometers). Daher ist die Beschreibung des zeitlichen Verlaufs nur so möglich, dass wir uns auf einen anderen Vorgang stützen" (in Macho, 1996, p. 161 recent edition; TLP originally published in 1922 (TLP 6.3611)).

[2] In response to M. Beaton's review comment, I want to add that whilst it is true that the time step in DST itself is

enactive approach, even if a mitigation of the shortcomings in computational systems by inclusion of an explicit clock and partial co-ordination is a half-blooded possibility (e.g., Clark, 1998; Cantwell-Smith, 1996).

The third point, to me, is the most obvious point and can even be argued against a dyed-in-the wool objectivist. Even if it were the case that time was basically Newtonian and even if there were no problems of synchronising the represented time in a Turing Machine with this real time, a simple fact is that *mental time does not work that way*. To take the most trivial example, everybody knows that in our experience, sometimes, time flies and sometimes, the hours go incredibly slowly. This is but one and one of the less interesting examples of how our mental time behaves strangely and at odds with Newtonian physics. The following section 8.2 is full of intriguing details about how mental time is structured and which embodied rules it follows, from phenomenology to psychophysics. These descriptions sometimes evoke the metaphor of a Newtonian-Cartesian time arrow, but only as a side issue in the fascinating and complex story of mental time.

Taking these three classes of problems together, they suggest one common thing: a Newtonian-Cartesian-cognitivist approach idealises away the real puzzles and mysteries of time cognition before even starting the scientific work. The classical computationalist modeller will end up wasting her time solving artificially induced technical problems resulting from the choice of formal language, but not address any of the really interesting questions, which are buck-passed to be solved by a homunculus working with the internal representation of external time. As an enactivist, I set out to find the meaningful sensorimotor invariants, given our natural habitat and evolutionary history, that lead us to construct our perception of time that stably and at all levels accompanies our mental activity. This approach seems infinitely more difficult, yet infinitely more satisfactory.

8.2 Time and its Many Dimensions in our Mind

This section attempts to represent in a string of words the complex space of evidence and ideas many able thinkers and scientists have had on time cognition and perception and how it is constructed and to relate them to each other. I chose to start off with merely phenomenological descriptions of time (subsection 8.2.1), referring predominantly to James' work that had been explicitly influenced by Husserl, as well as making reference to the work of Husserl, Merleau-Ponty and other 'real' phenomenologists. The following subsection 8.2.2 stays within the realm of conceptual contemplation, but focuses on those thinkers that explicitly relate the nature of time to physical processes, such as Kant and Piaget. The third subsection 8.2.3presents empirical approaches that rely in some form on written experiential reports, such as Núñez' anthropological work, Shanon's research on altered states of consciousness and Piaget's experiments in children's cognitive development. The fourth subsection 8.2.4 presents evidence from neuroscience and psychophysics, which makes direct reference to physical processes which may play a role in the constitution of time perception. Subsection 8.2.5 summarises and tries to bring these diverse perspectives together.

Before starting this journey, I would like to point out some issues that recur across authors

also arbitrary, endorsing embodiment and situatedness, it is automatically identical for both inside and outside, whereas a representationalist model the computer clock needs to be synchronised with the environment.

and disciplines, to prime the reader and ease the task of seeing the connections in this broad spectrum of work. Firstly, nearly all researchers that have seriously dealt with explaining time perception have remarked that *there is a primitive/intuitive temporal dimension inherent in our flow of consciousness and that it is different from our cognitive conception of time*. However, there is a multitude of ideas about the exact nature of either and how levels of sophistication are structured and relate. Secondly, *a spatial metaphor of time* seems absolutely indispensable to any analysis of time and is frequently explicitly pointed out. It seems that the question of how the conception of space and the conception of time relate is of crucial importance in the enactive approach to mind. Thirdly, a close look at the notions of *knowledge and time* reveals that they are intricately linked, in a story that includes also the concepts of *agency and possibility*. This last point is possibly the most obscure and tacit one, and I hope to be able to make some of it explicit along the way. I have to stress that my own position on time construction is still not fully developed, in this section I only phrase and explicate interesting questions. An attempt to hint at an answer to some of them is undertaken in chapter 12.

8.2.1 Phenomenology

The work on the phenomenology of time (presenting some of Husserl's, James' and Merleau-Ponty's, thoughts) that I briefly summarise here have, for my purposes, merits and demerits. In this short space, I want to point out some interesting observations that have repeatedly been made and that contradict our vulgar Cartesian intuition of time as the passage of punctual moments on a line. However, I want to conclude by uttering a tentative criticism of the direction some of these analyses have taken, which is quite a bold thing to do given that I have only superficially dealt with the material.

The most fundamental observation on the phenomenology of time perception is that the "cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time" (James, 1890). Were our flow of experience but a chaining of punctual moments, *our experience would change, but we could never experience any change*. The just-past is always still present, as is that which is about to come. This dynamics of 'retentions' and 'protentions' has been analysed and described in detail by Husserl (in Steiner, 1997, recent edition; the original sources on the phenomenology of time perception here cited were published 1893-1914) and it forms the starting point for all other observations on the phenomenology of time, as the present is the smallest unit of experienced time.

These observations make us realise that these extended chunks of present do not change continuously, flowing like a river, but discretely, abruptly switching their overlapping yet different meaningful content. "The discreteness is, however, merely due to the fact that our successive acts of recognition or apperception of what it is are discrete. The sensation is as continuous as any sensation can be." (James, 1890). So, from a continuous and changing flow of primitive sensation, we construct and chain moments of recognition of *meaningful*, in the most rudimentary form, percepts. These discrete and chained moments are not of arbitrary length, they are rather short: as James observes: "The durations we have practically most to deal with – minutes, hours, and days – have to be symbolically conceived, and constructed by mental addition, after the fashion of those extents of hundreds of miles and upward, which in the field of space are beyond the range of

most men's practical interests altogether." (James, 1890). This distinction between Husserl's three levels of time experience (the symbolic-narrative, the 'immanent flow' of meaningful moments, the primitive experience of change) is important with respect to phrasing research questions in time cognition, by making clear which of these layers are addressed and how. As I briefly explain in subsection 8.2.4, Varela applies his neurophenomenological approach to link these three levels of time sensation/perception to certain dynamical properties of the brain and neural interaction, focusing on explaining the second level, the immanent flow of time.

The symbolic mental addition of larger time spans introduces a fundamental and interesting issue: the apparent paradox of experienced pastness. Supposedly, at any moment in time, only the present is real, not the past (nor the future). The memory of the past, and the anticipation of the future, are manifestations of the past and the future in the present, a kind of 'trace' as Merleau-Ponty (2002, recent edition; French original published in 1945) calls it. The question then, as James (1890) pointed out, is: "But how do these things get their pastness?" A memory cannot *be* the past because the past does not exist anymore. If a memory, instead, was a retrieval of the original train of discrete chunks of subjective experience, it would feel as present as it did when it was lived. The memory wears the sign of the past-made-present, and what this pastness consists of is a mystery.

As Merleau-Ponty (2002, original in 1945) points out, it is our capacity to remember and expect and thereby, in a certain sense, change direction on the flow of consciousness, which allows us to think of time as time. Paradoxically, through this conceptualisation of time, it ceases to be temporal: "It is spatial, since its moments are spread out before thought" (Merleau-Ponty, 2002, p. 482; original in 1945). This statement may appear cryptic at this point but will become much clearer over the next two subsections.

The phenomenologists have observed many more interesting properties of time cognition. Among them, the distinction between future and past; rhythmicity in the primitive flow of time; meaning, intentionality and objects of time. Valuable though these contributions are, I want to second Varela's remark that "We still lack a phenomenology of internal time consciousness where the reductive gestures and the textural base of the experience figure explicitly and fully." (Varela, 1999). From the observations that a) the rudimentary flow of sensations cannot in itself lead to a consciousness of this flow and b) that a complex and multi-layered process of construction forms the basis of our conception of time, all the authors recited (Husserl, James and to a point Merleau-Ponty) focus on the very abstract, general and disembodied layers of our time consciousness, treating that which temporal experience across time scales, modalities, tasks, behaviours, levels of abstraction has in common. As Varela remarks, Husserl's prime example of listening to a melody is even developed without us knowing if the melody is familiar, which kind of emotional effect it has, where and how it is heard.[3]

I think all the issues I just summarised figure in trying to investigate the sensorimotor origins

[3]Heidegger's *Being and Time* (1963, original in 1927) seems to be very much the exception to this rule; pages and pages are dedicated to the description and analysis of the temporality of different aspects of life, and reading through some of them, they seem very rich of perceptive and intelligent observations that are much more particular and to the point than those I just lamented to be too general. However, I have not fully read this piece of work, which I believe has to be understood as a whole. Therefore, I think I would not do justice to Heidegger if I sprinkled in some citations at this point to represent his view, and instead acknowledge that there is a very important and exceptional thinker whose work is missing from my analysis but should not.

of our perception of simultaneity by investigating experiments with delays - *how*, I develop later on. At this point, however, I think that the focus of the cited phenomenological reductions is biased towards the very particular experiences of time as an object of armchair contemplation, not time as an aspect of common or garden experience of any kind. Varela also lamented this in his neurophenomenological account of the granularity of time, yet he succeeded in linking concrete experimental results to these general and high level phenomenological descriptions. I hope to be able to do something similar with the study of sensory delays and simultaneity.

8.2.2 The Construction of Time

The psychological/phenomenological accounts summarised in the previous subsection deliberately discard reference to physical time, or reduce it to a minimum, in order to show how the idea of a simple four dimensional Newtonian-Cartesian time-space (a box plus an arrow) contrasts our experience of time on all levels. However, even if we cannot experience Newtonian time, as a tool, the concept and its geometrical and mathematical properties are very powerful in explaining, understanding and predicting the world around us. In this section, I present Kant's and Piaget's ideas on the question of how we come to think that this is how the world is and to fall for the Cartesian illusion that this is how we experience it as well.

To start with the discussion of time in Kant's *Critique of Pure Reason* (1974, original in 1787), in his transcendental aesthetics, Kant assigns a special status to time and space, calling them the *a priori* formal condition of *Anschauung* (perception). In an at least proto-constructivist fashion, he stresses again and again that time and space are not objectively real, in the sense that they are not observer-independent properties of the *Welt an sich* (world in itself). Time is nothing but the form of our inner senses, how we appear to ourselves.[4] As such, time has 'empirical reality', 'subjective reality' for Kant, and it makes the perception/imagination of myself and my subjective experience of my own subjectivity as an object possible.[5] So far, this observation resonates strongly with the phenomenologists' identification of the primitive and immanent levels of time experience, even if Husserl points out that reflexive experience of change as change is itself atemporal and thus not part of the immanent flow of time (in Steiner, 1997, p. 327; (original cited as ca. 1909)). Furthermore, the diagnosis of temporality of direct subjective experience as a necessity for the experience of self is very much related to Heidegger's (1963, original in 1927) idea that temporality is necessary for concernful existence.

This irreducible reality of subjective time *a priori* in itself does not, however, contain the categorical and relational properties that characterise our grown adult conception of time. Time is not part of the experienced exterior, is not a property of gestalt, location, etc, but determines the relation of experience in our inner state. The lack of gestalt of our inner state is compensated for by the construction of a metaphor such as time as an arrow that goes to infinity, chaining 'manifolds'.[6] Thereby, time is assigned to the world and becomes a property not just of myself

[4] "Die Zeit ist nichts anders, als die Form des innern Sinnes, d.i. des Anschauens unserer selbs und unsers innern Zustandes" (Kant, 1974, p. 80f; original in 1787).

[5] "[Die Zeit] hat also subjektive Realität in Ansehung der innern Erfahrung, h. i. ich habe wirklich die Vorstellung von der Zeit und *meinen* Bestimmungen in ihr. Sie ist also wirklich nicht als Objekt sondern als die Vorstellungsart meiner selbst als Objekts anzusehen" (Kant, 1974, p. 83; original in 1787).

[6] "Denn die Zeit kann keine Bestimmung äußerer Erscheinungen sein; sie gehöret weder zu einer Gestalt, oder Lage, etc., dagegen bestimmt sie das Verhältnis der Vorstellung in unserm innern Zustande. Und, eben weil diese Innre

but of objects around me. Crucially, Kant sees the construction of time as an object or a dimension of the objective world as a strictly logical process: apart from 'empirical reality', time possesses 'transcendental ideality'. He supports his claim with the fact that mathematical and logical laws hold for time and space, which are strictly intersubjectively valid and thus not really *a posteriori* either (synthetic judgments *a priori*).

Kant's observations point out some very valuable issues: that temporal and spatial experience in its rudimentary form cannot be stripped off our experience and imagined away in the way that other aesthetic qualities, such as hardness or colour can. Furthermore, that temporal experience is tied even closer into experience than spatial experience: space is a property of the exterior, whilst subjectivity is experienced non-spatially, yet temporally - remark that Cartesian fantasies of brains in vats or the matrix are happy to place the *res cogitans* in an illusory fantasy world as regards its spatial surroundings (with other spatial surroundings); the time line, however, in which the deception takes place is real, because it is the *a priori* form of the subject. What I want to debate, however, is the privileged character that Kant assigns to the constructed objective 'transcendentally ideal' time: I think that the elaborate observations by the phenomenologists, as well as the empirical data presented in the following sections 8.2.3 and 8.2.4 shows that there are many and variable factors contributing to the conception of time (culture, sensorimotor dynamics, development, intact functioning of the brain, etc.) and that the logical properties are, to a degree, contingent on these factors. The empirical study of the construction of time shows that our temporal thinking is more than just *a priori* primitive time and logic, because there are experiences of time as an abstract concept that violate logical constraints and lack consistency.

I want to compare Kant's ideas about time with those of Piaget, who distinguishes *intuitive time* and *operational time*. Intuitive time for Piaget is "limited to successions and durations given by direct perception." (Piaget, 1969, p. 2; original in 1946), which seems to broadly correspond to what Kant describes as *a priori* "Ansehung der innern Erfahrung' (observation of inner experience) (Kant, 1974, p. 83; original in 1787), whilst operational time "is the operational co-ordination of the motions themselves" (Piaget, 1969, p. 3; original in 1946) and builds on the active successive construction of the relations between simultaneity, succession and duration.

In many ways, Kant's and Piaget's views are very similar: both distinguish two and only two modi of time, the primitive and the constructed (intuitive vs. operational in Piaget, empirical vs. transcendental in Kant). In this sense, both simplify the multiplicity of layers and dimensions in temporal experience that the phenomenologists recognise. However, maybe as a consequence of this simplified view, both give in to the urge to want to explain the basis of the construction of the more complex conception of time from the primitive concept and the concept of space. Piaget hypothesises a previous construction of space and geometrical relationships as the basis of the development of a more sophisticated concept of time: "It is only once [space] has already been constructed, that time can be conceived as an independent system" (Piaget, 1969, p. 2; original in 1946). In a for me not entirely transparent way, Piaget seems to assign ontological priority to the conception of space ("space is above all a system of concrete operations, inseparable from the experiences to which they give rise and which they transform" (Piaget, 1969, p. 1; original

Anschauung keine Gestalt gibt, suchen wir auch diesen Mangel durch Analogien zu ersetzen, und stellen die Zeitfolge durch eine ins Unendliche fortgehende Linie vor, in welcher das Mannigfaltige eine Reihe ausmacht" (Kant, 1974, p. 80f; original in 1787).

in 1946)) over the concept of time ("In the course of its construction, time remains a simple dimension inseparable from space" (Piaget, 1969, p. 2; original in 1946), whereas "space suffices for the co-ordination of simultaneous positions but as soon as displacements are introduced they bring in their train distinct and therefore successive spatial states whose co-ordination is nothing other than time itself." (Piaget, 1969, p. 2; original in 1946)).

Further on in the Critique of Pure Reason, Kant also hints towards some of the relations between time and space and their geometrical properties that he thinks form the basis for the synthetic judgments *a priori* that constitute transcendentally ideal concepts of time and space. In the analogies of experience (transcendental analytics), Kant explains how the concepts of constancy, succession and simultaneity (*Beharrlichkeit, Folge und Zugleichsein*) result from connecting distinct experiences in subjective time. For instance, he points out that simultaneity in time is given if the order in which objects are perceived is arbitrary or reversible, for if the order in which I experience them was fixed, they would be successive and not simultaneous.[7] At the same time, he asserts that the rules of constancy, succession and simultaneity are *a priori* and necessary for experience to happen at all.[8] Interestingly, the identification of reversibility as characteristic of space also gives additional significance to Merleau-Ponty's (2002, original in 1945) observation that the possibility to anticipate and remember, which allows us to travel freely in both directions on time's arrow and to thus objectify it, corresponds to a spatialisation of time. According to Kant, only when things other than myself move around, the conceptions of time and space get in contact, start to resemble each other and require the construction of relations such as movement, velocity/speed, simultaneity and causality in order to distinguish them and disambiguate, which results from the processes and experiences described by Kant in his analogies of experience.

Both Piaget and Kant, in my opinion, have rushed over a number of steps along the way of how temporal experience is constructed. In my opinion, Kant has focussed too much on the logical-mathematical side of space and time, whilst Piaget has acknowledged the gradual development of these conceptions and the role of sensorimotor experience in the course of development. Piaget, on the other hand, fails to address the complexity and reciprocity of the steps that lead towards the construction of sophisticated concepts of both time and space. Kant analyses from primitive temporal experience (i.e., the experience of change) and primitive spatial experience (i.e., the experience of inside/outside), more operational conceptions of both time and space are bootstrapped, in a process of gradual and stepwise co-construction. This gradual and stepwise co-construction has been recognised by Kant, even if he focuses too much on the disembodied cognitive reasoning and logics.

Nonetheless, Piaget and Kant have both addressed the construction of time from movement and in relation to space, which helps to ask the question of the embodied origin of the experi-

[7] "und darum weil die Wahrnehmungen dieser Gegenstände einander wechselseitig folgen können, sage ich, sie existieren zugleich" (Kant, 1974, p. 242; original in 1787) or later "Woran erkennt man aber: daß sie in einer und derselben Zeit sind? Wenn die Ordnung in der Synthesis der Apprehensionen dieses Mannigfaltigen gleichgültig ist, d.i. von A, durch B, C, E, auf E, oder auch umgekehrt von E zu A gehen kann. Denn, wäre sie in der Zeit nach einander (in der Ordnung, die von A anhebt, und in E endigt), so ist es unmöglich, die Apprehension in der Wahrnehmung von E anzuheben, um rückwärts zu A fortzugehen, weil A zur vergangenen Zeit gehört, und also kein Gegenstand der Apprehension mehr sein kann" (Kant, 1974, p. 243; original in 1787). Thanks are due to C. Lenay (2003) for pointing me to this part of the Critique.

[8] "Daher werden drei Regeln aller Zeitverhältnisse der Erscheinungen, wornach jeder ihr Dasein in aller Erfahrung vorangehen, und diese allererst möglich machen" (Kant, 1974, p. 217; original in 1787).

ence so elaborately analysed and described in the phenomenological investigations summarised in the previous section. Taking together these theoretical accounts, the phenomenological and the constructivist, the question that emerges is: what is it that makes the temporal experience in all its levels of subtlety, complexity, symbolicity, etc. such stable outcomes of a developmental process? Remembering Jonas' (1966) and Weber's (2003) arguments about how our own experience as living organisms also helps us to understand meaning and value in the animal kingdom, the behavioural dimension of the concepts of time and space can be transferred and applied to animal cognition[9] and, by comparing species, cultures, developmental stages, pathological experience of time, etc., it will be possible to come to a unified account of time and space cognition.

8.2.3 Studies on Human Time Cognition Based on Language and Verbal Reports

In this subsection, I analyse some of the empirical evidence on human behaviour and subjective reports from different disciplines and how it relates to the theoretical contemplations presented so far. I want to start by briefly summarising some of Piaget's experimental results from his work with children. Then I address some anthropological and linguistic work, and finish with a summary of Shanon's work on temporal experience under the influence of the psychedelic Ayahuasca potion.

Out of the experiments about the construction of the child's conception of time, I find those about succession and simultaneity in physical time most revealing.[10] The experimental paradigm used in both cases is the simultaneous motion of two figures at different velocities, either stopping simultaneously or successively. As a consequence of the difference in velocity, these events can lead to different spatial configurations once both have stopped. This leads to interesting confusions in children at certain stages of development as when they are asked about spatial displacements, temporal orderings and how these two relate. In what Piaget calls 'stage I' "successions and durations remain undifferentiated from distances [...] and differences in speed are thought to preclude synchronous processes and lead to confused estimates of duration." (Piaget, 1969, p. 85; original in 1946). It is worthwhile to give some of the full text of an experiment with a four year old child in order to get an impression of the kind of errors children at stage I commit. The child is presented with a situation in which a yellow figure is made to stop earlier than a blue figure, with the blue figure still stopping spatially less far than the yellow figure (child's responses in italics):

> "Did they stop at the same time? *No.* Which one stopped first? *The blue one.* Which moved longer? *The yellow one.* [...] But which one stopped first? *The yellow one. No, it was the blue one, the yellow one went on longer.* Let's do it again. (The race is re-run.) *The yellow one stopped first, the blue one was still moving, so the yellow one went on longer.* But did one stop before the other? *The blue one*" (Piaget, 1969, p. 86; original in 1946).

Children at this developmental stage are incapable of detecting or correcting their confusion of temporal and spatial differences, and do not seem to be bothered by the logical contradictions either. To pre-empt objections that there could be just a linguistic confusion about the spatial metaphor in time, it should be mentioned that Piaget colleagues asked the children further less

[9] For instance, as a consequence of Kant's observations about reversibility and space, I believe that the world of e-coli bacteria is non-spatial - a theory that will have to be developed and well argued.

[10] The experimental paradigm used in the analysis of the construction of duration and sequencing in elementary time, which Piaget presumes to be more basic, is, in my opinion, more difficult to interpret.

ambiguous and more intuitive questions, coming to the same results of the children confusing temporal and spatial order. More interestingly, maybe, the systematic mistakes disappear, at the same developmental stage, if the figures are made to move into opposing directions. Piaget describes how children pass through later developmental stages in which they would still make misjudgments of the described type, but be able to correct them when being pointed, in dialogue, to the logical contradictions in their report, before finally arriving at a 'transcendentally ideal', in Kant's sense, conception of time. It may be worthwhile to remark, as well, that these stages and mistakes are broadly the same if the child itself is made to run against the experimenter, rather than to just observe the figures.

What do these experiments show us? They exemplify that there can be perception of time which goes beyond the mere recognition of change, which we identified as primordial flow of consciousness and condition for subjectivity, but which still is not a full-blown transcendentally ideal space which can persist and be viably applied during several years of child development. The children have clearly learned to assign temporal properties to objects in the world and that there are order relations for events such as stops, but these remain fuzzy and intermingled with spatial properties. The mathematical and logical properties of time are not yet fully developed, i.e., there is no clear distinction between changes in a previously registered flow of consciousness that are really reversible (spatial) and those that are only mentally reversible (temporal).

I now want to introduce some interesting findings about the use of temporal language across culture, starting off with Lakoff and Johnson's (2003) research of how spatial language is used metaphorically to talk about time in nearly all languages: the future is seen as being in front whereas the past is conceived of as behind in expressions such as 'The time will come when . . . ' or 'In the weeks ahead of us' (Lakoff & Johnson, 2003, p. 42). Lakoff and Johnson's work shows that such 'conceptual metaphors' are used systematically and consistently, demonstrating semantic links, not just verbal shorthands, and that such conceptual metaphors are abundant across cultures and contexts. Interestingly, Lakoff and Johnson identify some apparent inter- and intra-cultural inconsistencies in the spatial metaphor of time which can be resolved by introducing agency into the metaphor, i.e., to conceive time passing as motion, which can be instantiated either as time being the moving object or us as moving in time (Lakoff & Johnson, 2003, p. 41-45).

A very interesting deviation from the described conceptual metaphor has been described by Núñez and Sweetser (2006) to occur in the Aymara language spoken by indigenous people in some parts of the Andes. In a crude simplification of Núñez and Sweetser's findings, the Aymara language is to date the only reported language in which the *time is space* metaphor is directionally inverted (i.e., the past is in front of the speaker and the future lies behind the speaker). Most intriguingly, Núñez and Sweetser have also found that the accompanying gestures of the Aymara speakers comply with this use of language (e.g., an Aymara speaker would would point forwards when using the Aymara word for forward and when referring to the past) and that this seeming spatial inversion of temporal gestures is preserved when native Aymara speakers speak the Andes dialect of Spanish and partially adapt Spanish grammar to match their metaphorical purposes.

The *time is space* metaphor described by Lakoff and Johnson seems so naturally linked to the processes of spatialisation and temporalisation through embodied experience of the world we have discovered so far. How is it possible that the Aymaran people use the metaphor in its

inverse direction, contradicting all our intuitions about how time is metaphorically spatialised? Núñez and Sweetser have a very interesting explanation for this exceptional use of the *time is space* conceptual metaphor in language and gesture: they observed that Aymaran spatial metaphors for time never involve any self-motion. Whilst the *time is space* metaphor in most languages involves movement along a path or a river (either by the subject or by an agent-time itself), leaving behind visited (past and known) stations and discovering the new behind the next corner, the spatial metaphor of time for the Aymaran people is a static one, in which the space in front of the subject is visible and thus known, whilst the space behind the speaker is unknown and changes occurring behind the speakers back can go undetected and surprisingly. Interestingly, the authors also point out that the Aymaran culture assigns importance to personal testimony and that they discredit talking about the speculative/unknown, which is marked by a reluctance to talk about the future in general.

These interesting findings show two things: firstly, it is impossible to talk about *the* spatial metaphor of time. There are variable structural similarities at different levels of meaning and interpretation between the two. The common *time is space* metaphor actually has to be elaborated to be a *time is motion along a path* metaphor, whilst the metaphor of *the past is known/visible and the future is unknown/invisible* can lead to an alternative interpretation of how location corresponds to a conceptualisation of time, a more passive and backward looking one. Secondly, it emphasises an aspect of temporal conception that I have not yet covered very much, i.e., what distinguishes the past from the future among the temporal constructs that are not the present. The past does not change, neither by its own accounts, nor by my influence, whilst the future is open and not yet fixed. This relates to Merleau-Ponty's remark that the future seems to only exist "by analogy" (Merleau-Ponty, 2002, p. 481; original in 1945), guessing that this moment will pass and turn into past like all the ones before it, being replaced by another one that is yet unknown.

In my opinion, the example of the *time is space* metaphor in the Aymara language is an important jigsaw piece in putting together the known and the unknown, the real and the possible, action and passivity, subject and object to characterise some fundamentals of temporal and spatial experience and how they relate. Having identified the experience of change and the experience of inside/outside as those components of experience that are truly irreducible and that seem to form the foundations of the constructed notions of time and space we can now identify past experience as that which is unchangeable and known and future experience as that which is changeable and unknown. Spatial relationships are those that characterise the systematicities with which endogenous movement influences the flow of conscious experience, both in the past and in the future. In a static world, time and space would be indistinguishable because the one would fully determine the other, all would exist simultaneously. It is only when exogenous action comes in that the future becomes indeterminable, that the past experience contains non-spatial components which have nothing to do with self-motion. Motion and velocity of external objects are constructed.

As a third lesson from Núñez' results, we should be gently reminded that our conception of time is not just contingent on developmental phase, that it is not only phenomenologically more complex and multi-faceted than Kant seems to acknowledge, but that, in its complex structure, temporal experience will also have a strong cultural component. Evans (2004) lists nine different (yet related) meanings of the word/concept 'time' in English, four of which he claims to be

"secondary lexical concepts", i.e., they are cultural constructs, rather than rooted in universal human experience. Even though it is not clear to me how exactly he draws this line, it seems more than plausible that the construction of time as a general concept and dimension of all things, even though possibly departing from intuitive/primitive/empirically real/immanent time of subjective experience per se, will, in its details, be culturally contingent. The Aymara culture with its passive and backward looking attitude towards time is maybe the best example for this cultural component.

As a last source of data on contingency in the constructed perception of time, I want to briefly summarise Shanon's (2001) research on temporal perception under the influence of the psychedelic potion Ayahuasca. There are a number of alterations of temporal experience induced in both natives of the Amazon forest from those cultures in which Ayahuasca is traditionally used in a ritual context, and naïve European and North American participants. Shanon describes some rather gentle alterations of temporal experience (change of rate of experienced time, change in perceived distance to past or future events, relocation of 'present' in the illusion to witness/live past events). These examples are interesting, because they point out aspects of temporal experience, which can be altered and distorted whilst leaving the general concept of time intact, and which I have not yet addressed at all. The construction of pastness, futureness and presentness and the order relations in time seems more pressing, but the quantitative properties of time are as fundamental and inherent in the experience of the symbolically constructed time scales as the qualitative ones are.[11]

More related to the previous analysis of the nature of the concept of time and space are, however, experiences under the influence of Ayahuasca that induce the feeling of timelessness, eternity and the confusion of perception, memory and anticipation. The latter is the temporal variant of a more general effect of Ayahuasca that the real and the unreal get blurred. In the case of time, however, this blurring is interestingly related to our previous analysis that reality plays an important role in the constitution of pastness, presentness and futureness. In the limit case, the blurring of these boundaries results in states of consciousness which I would want to call the *completion of the time is space metaphor*. As Shanon puts it, "the temporal may, in a fashion, be reduced to the spatial" (Shanon, 2001, p. 47). To quote a report from such a vision:

> "In front of me I saw the space of all possibilities. The possibilities were there like objects in physical space. Choosing, I realized, is tantamount to the taking of a particular path in this space. It does not, however, consist in the generation of intrinsically new states of affairs" (Shanon, 2001, p. 47).

Interestingly, Shanon reports that such an 'out of time' experience is frequently accompanied with the feeling of omniscience, stripping the future of its speculative and open character. Resonating with Heidegger's (1963, original in 1927) ideas of temporality being the basis for concernful existence, the stepping out of time coincides with a loss of concern, temporality becomes irrelevant, a side effect "is the taking of things less seriously and with more tolerance, forgiveness and also a (benevolent) sense of humour" (Shanon, 2001). This experience of eternity and the complete spatialisation of time is a perfect instantiation of what Husserl describes as god's consciousness, a "limit-notion of temporal analysis: god's infinite consciousness contains all times at once. This

[11] The reader is referred to James' (1890) contemplations of clocks, rhythmic experience and counting on these matters and to Piaget's (1969, original in 1946) investigation of the construction of duration in temporal experience.

infinite consciousness is a-temporal" (in Steiner, 1997, p. 40; original cited as 1893-1917).[12] And just as Husserl realises that "even a divine consciousness would have to progress temporally" (in Steiner, 1997, p. 40; original cited as 1908-1914)[13], the description of experienced eternity under Ayahuasca is constrained in the same way: "Further, it should be noted that while traveling in the space of possibilities takes time, the possibilities themselves are there, given in an ever-present atemporal space." (Shanon, 2001, p. 47). In this experience of being outside time, all agency but my own disappears, and thereby all uncertainty. Time is spatialised and loses its meaning. However, even as the constructed notion of time in many of its dimensions collapses, the primitive flow of time *a priori* persists.

These reports from various types of temporal experiences of the constructed type that seem alien to the healthy sober adult westerner help to illustrate some of the constitutive qualities and regularities of temporal perception and how time relates to space, knowledge, subjectivity and concern. They also illustrate which of the qualities of everyday temporal experience are contingent and can, even if at the cost of logical consistency, self-concern or the notion of time itself, disappear. Thereby, they give us an impression of what is left: contemplating all these examples, we may get a better idea of the absolute flow of consciousness that survives all these bizarre transformations and seems indeed necessarily linked to all human experience. Similarly, as the symbolic level of time experience is constructed from the primitive and immanent flow of time, it is constrained by the latter. Investigating these constraints may reflect back on the study of the former.

8.2.4 The Brain, Sensorimotor Dynamics and Primitive Time Perception

The previous section has given us an impression of how the complex of temporal experience of symbolical time can be variably constructed *on top of* the immanent flow of conscious experience, the primitive registration of change. This analysis, though interesting in itself, serves more to *separate* some aspects from what we mistakenly and intuitively conceive of as a unified irreducible temporal experience. In order to empirically investigate the more primitive levels of time consciousness, i.e., how the immanent flow of temporal object-events is constructed from and relates to the primitive *a priori* flow of consciousness, and how these two layers interact with the symbolic conceptualised experience of time, as the phenomenologists describe it (subsection 8.2.1), we have to access physics directly and surpass human consciousness and socio-linguistic self to a certain degree. In this subsection, I present empirical work investigating these more primitive layers of time experience.

An important project in the endeavour to explain the immanent flow of time with reference to the brain and the body is Varela's 'The specious present: a Neurophenomenology of time consciousness' (Varela, 1999), in which he links the three levels of temporal experience identified by Husserl to dynamical properties of the human brain.

Even though Husserl plausibly argues *that* there are three levels of time experience, i.e., 1.) the primitive and continous flow of sensations, 2.) the discrete chaining of meaningful 'nows' and

[12] My translation: "...Limes-Begriff der Zeitanalysen: 'Gottes unendliches Bewußtsein umfaßt alle Zeit 'zugleich'. Dieses Bewußtsein ist unzeitlich." (in Steiner, 1997, p. 40; original cited as 1893-1917).

[13] My translation: "Selbst ein göttliches Bewußtsein müßte notwendig zeitförmig verlaufen" (in Steiner, 1997, p. 40; original cited as 1908-1914).

3.) the symbolically constructed narrative time level that exceeds in duration our experience of the present (see section 8.2.1), he does not give any reasons why that should be the case - why not just one level? Why not infinitely many?

As part of the collection 'Naturalizing Phenomenology' (Petitot, Varela, Pachoud, & Roy, 1999), Varela attempts to fill this gap. Quantifying the temporal duration of changes in each level, the continuous flow of sensations is identified with the duration of several tens of milliseconds. This is the time scale in which we humans can make minimal perceptual discriminations (even if exact resolution varies across modalities) and the time scale of inter-neural events (action potentials). It is not necessary to be a representationalist to recognise the essential role the brain plays for the human mind and behaviour, and, as a bottleneck, the physiological limitations of the brain surely impact on the construction of the meaningful world. Changes that happen faster than the fastest meaningful processes in the brain and body, for our purposes, simply do not exist.

From this time scale of minimally perceptible change, the second level of time consciousness, Husserl's 'immanent flow of time', is constructed. According to Varela, the time scale of this level is in the scale of around 1s (and is thus in the same ballpark as James' (1890) 'specious present' of 3s), a time scale which corresponds to the time necessary to integrate several of the atomic sensations identified as the units of the primitive flow of consciousness. This is the level of recognised change, the level in which experience becomes subjective and present, in a very rudimentary form meaningful. Therefore, the immanent level of time consciousness is also called the level of 'temporal object-events' (Husserl's *Zeitobjekte*, Varela, 1999), because perception of present has a meaningful content, even if those are not properly reflexively and objectified and transcendental objects or eventified transcendental events. In his neurophenomenological approach, Varela focuses on explaining the construction and delimitation of the second 'immanent' level from the first 'primitive flow' level and does not attempt to directly naturalise the third level, that which James identifies as the level that is symbolically constructed, and which Varela calls the scale of "descriptive-narrative assessments" (Varela, 1999). Symbolic cognitive processes are very difficult to link to physical or physiological processes - little is known about even the most basic human use and construction of symbols, as pointed out in section 2.4. The studies based on verbal reports presented in the previous sub-section 8.2.3 appear at this moment much better suited to elucidate the mysteries of the objectified, symbolised and abstract experience of time.

Varela's identification of neurodynamic processes at qualitatively different time scales, in which, naturally, qualitatively different processes of variable exact duration happen is interesting because it naturally accounts for the variability in length of the chunks that constitute each level. It is tempting to identify a 'magic number' of neural meaning: for instance, Libet (2004) came across the time span of 500 ms repeatedly in his work, equating it with the fundamental unit of time cognition, and a similar thing happened to Michael Herzog (personal communication, Nov. 2007) with the time span of 300 ms. However, events of any duration can be possibly meaningful, within their temporal realm, and the neurophenomenological approach gives physical reasons for why this should be so. This also implies that there are time spans which are at the overlap of time scales, a fact that I return to in chapter 12.

An interesting piece of evidence with respect to the physiological basis of present-time consciousness is also Libet's (2004) classical neuroscientific work on measuring neurosensory and

neuromotor latencies and their relation to experienced simultaneity. By means of cortical stimulation of variable length, Libet found through his experiments in cerebral stimulation in the late 1950s that a stimulus was only registered and reported if it persisted for 500 ms, which Libet found to be "suprisingly long for a neural function" (Libet, 2004, p. 39). This led him to the conclusion that "*awareness of our sensory world is substantially delayed* from its actual occurrence" and that we are thus "always a little bit late" (Libet, 2004, p. 70). Libet found a delay of nearly identical length to pass between the neural potential ('Readiness Potential') marking a 'point of no return ' in decision making and the awareness of making this decision, which subjects indicated with reference to a clock, and which is the research he became most famous for. These very straightforward experiments pose a challenge to our intuition that experienced time is coordinated and synchronised with the 'objective physical time' as it is measured by a clock.

Moreover, Libet's experiments show how, by means of measuring and controlling perceptual judgments and correlated physical processes, links can be established between the physical (in this case neurophysiological) processes and experience (in this case the 'middle' immanent flow of experience and temporal object-events). As already remarked in chapter 3, section 3.5, Libet's way of accessing experience, i.e., via perceptual judgments, is in line with the psychophysics measurements of experience laid out by Fechner (1966, recent edition; German original published in 1860) and thus, in my opinion, implements what Fechner envisioned as 'internal psychophysics' and which was not possible at his time because of technological limitations to the study of the brain.

As concerns the more traditional discipline of 'external psychophysics', there are also a number of interesting and related results, some of which I want to quickly summarise. One interesting phenomenon studied in psychophysics is the flash-lag-effect (FLE, e.g., Nijhawan, 1994): if subjects are presented with a moving bar, half of which is constantly illuminated and half of which is flashing, the flashing part of a moving bar appears to lag behind a constantly illuminated part, with the spatial distance being a function of the velocity of the bar. Nijhawan postulates that this difference in perceived location is due to the "predictability of the continuous segment and the unpredictability of the strobed segments" (Nijhawan, 1994, p. 257), such that "the perceived location [. . .] is closer to the object's physical location than might be expected from neurophysiological estimates" (Nijhawan, 1994, p. 257) to make real time interaction possible.

Two other phenomena studied in psychophysics I want to briefly address are 'backward masking' (or 'retroactive masking') and apparent motion. In backward masking, a stimulus (peripheral (e.g., Herzog, 2007) or cerebral (e.g., Libet, 2004)) administered to an experimental participant sometimes suppresses the awareness of a previously administered stimulus, even if the participant had become aware of the previous stimulus had the second masking stimulus not been presented. This retroactive interference of a later stimulus with awareness of a previous one shows that present-time experience of a stimulus is indeed contingent on what happens within a time window after the presentation of the stimulus. Similarly, in apparent motion (e.g., Gepshtein & Kubovy, 2007)[14], two discrete subsequent and displaced presentations of visual stimuli are perceived as a continuous motion from one to the other. Therefore, continuously experienced motion is contingent on the presentation of the second stimulus, which only marks the end point of the

[14]Thanks are due to O. Gapenne for pointing my attention to this phenomenon.

experienced motion.

Such results from psychophysics and from Libet's neuroscientific work serve to point out in how far our observer perspective on physical time and our observer perspective on another subject's temporal experience differ and in how far they relate. These systematic discrepancies even on the more fundamental level of the immanent flow of time can be studied and can help to explain the construction of more fundamental time cognition, on a level that is left invariant under higher level transformation (e.g., cultural or drug-induced variation, cf. section 8.2.3).

Recognition, interpretation and explanation of these systematic discrepancies, however, can take different forms depending on the choice of paradigm. From an objectivist-representationalist stance, these discrepancies can be seen as imperfections in the brain's representation of Newtonian physical time. Such a perspective clearly impacts on the questions these intriguing phenomena raise. For instance, Libet observes: "so we have a strange paradox: Neural activity requirements in the brain indicate that the experience or awareness of a skin stimulus cannot appear until after some 500 ms, yet subjectively we believe it was experienced without such a delay" (Libet, 2004, p. 72). Libet resolves this seeming paradox of evident synchronisation with the world (first person experience and successful real-time behaviour) and, at the same time, apparent lagging behind the world (third person observations of experienced time vs. physical time) by proposing mechanisms that backdate experience to the time of their 'real occurrence'.

Similarly, even on this fundamental level of time perception, the questions of time and space are already, or possibly even more intertwined. An objectivist-representationalist perspective on this intertwinement sees it as a problem and strives to resolve it and take it apart. Eagleman and Sejnowski (2002), for instance, argue that the FLE may have been misconceivably considered a temporal illusion. They argue that the effect could instead be a spatial illusion, resulting from inaccuracies in the inference processes that the brain performs to determine the location of the constantly illuminated part of the bar and the temporal cost of performing this computation. Immaterial of the evidence they base their argument on, I believe that the objective of untangling spatial and temporal illusions at this level of experience is a bit misled because, from a constructivist perspective, the kind of phenomena under investigation form the very basis for distinguishing time and space. Therefore, there is the possibility for phenomena, perceptions or behaviours where such a distinction is *impossible*. Trying to impose it distracts from the real questions at stake, which is: what are the origins of our conception of time and space? This question is turned upside down when investigating the artificial problems of how external time/space can be internally represented.

From a constructivist perspective, no coordination other than that of real physical behaviour in the real physical world is necessary. Nijhawan (2004) has recently contradicted his own previous hypothesis that the FLE is consequent to a neural delay compensation mechanism inferring an objects' 'real' position (1994). He argues that "the 'real' in the ['*vdt*-lag' premise] is an unobservable quantity" because, in closed loop interaction, "many features of 'real' objects 'out there' (e.g., position) are due to descending (internal) neural signals, processes that are related to feed-forward motor control and to Helmholtz's notion of re-afference. The view that emerges is that an output of one modality (e.g., object-position given by the visual system) can be related (compared) to the output of another modality (e.g., hand-position given by the motor system), but not to some ideal-

istic 'really' given position." (Nijhawan, 2004).[15] Importance of closed loop interaction predicts a similar flash lag effect to appear in motion, which has been empirically confirmed (Nijhawan & Kirschfeld, 2003). The proto-constructivist insight just sketched informs the multilayered and integrated theory of visual prediction recently expressed in (Nijhawan, 2008).[16]

A more general characterisation of the intertwinement of space and time in embodied and situated perception is phrased by Gibson in his ecological perception approach. Gibson postulates that "we have accepted space-perception as a valid problem, but have been uncomfortable about time-perception. We have attempted to keep separate the problem of detecting patterns (objects) and that of detecting sequences (events). And hence the equivalence of pattern and sequence, of space and time, has seemed to be a puzzle which had better be swept under the rug than confronted" (Gibson, 1982, p. 174; recent edition; original published in 1966). Taking into consideration the sensory physiology of humans, Gibson characterises the situation as follows

> "The eyes of primates and men work by scanning - that is, by pointing the foveas at the parts of a scene in succession. The eyes of rabbits and horses do not, for they see nearly all the way around at once and have retinas with little foveation. Does this mean that a horse can perceive his environment, whereas a man can apprehend it only with the aid of memory? I once thought so on the theory that successive retinal images must be integrated by memory, but this now seems to be wrong. It is truer to suppose that a visual system can substitute sequential vision for panoramic vision, time for space. Looking around is equivalent to seeing around, with the added advantage of being able to look closely. It is no harder for a brain to integrate a temporal arrangement than a spatial arrangement" (Gibson, 1982, p. 174; original in 1966).

Gibson's insight and his conclusion that "the perception of space is incomprehensible unless we tackle it as the problem of space-time" (Gibson, 1982, p. 175; original in 1966) resonates with Lenay's assessment that "if perception is constituted at the core of a closed sensorimotor loop, enriching perception [...] should be equally possible by means of enriching the sensory inputs at any moment or by means of enriching the repertoire of possible actions" (Lenay, 2003, p. 57)[17], which has been investigated by means of experimentation with receptive field parallelism.

The perspectives just outlined, i.e., Lenay's, Nijhawan's and Gibson's, as well as Varela's neurophenomenological perspective of course, even though not all of them may be based on a strictly constructivist epistemology, are constructivist in that the account of time or space perception given does not already contain as an explanatory premise. No Newtonian concepts of time or space are presumed *a priori* as target outcomes for processes of internal representation. This departure from the objectivist premise is intricately linked to the recognition of the importance of closed loop sensorimotor interaction. Taking such a constructivist view on time and space perception is a liberation from the apparent paradoxes of coordinating internal time and external time or

[15]This position also has been argued and expanded in a talk in the DyStURB reading group at the University of Sussex in March 2005, where Nijhawan elaborated on the logical complication that the act of temporal measurement by the observer-scientist poses (personal communication).

[16]It has to be pointed out that the criticism of Libet's interpretation of his own results here presented resonates strongly with the ideas expressed by Mikael Karlsson in an oral presentation during the Compiègne Seminar in Jan. 2007 and in an unpublished draft paper called 'Perception, Interaction, Time'.

[17]My translation: "En effet, si la perception se constitue au cœur du couplage sensorimoteur, elle doit pouvoir être enrichie [...] aussi bien par un enrichissement de l'entrée sensorielle délivrée à chaque instant, que par un enrichissement du répertoire des actions possibles" (Lenay, 2003, p. 57).

those resulting from an attempt to see the distinction between space and time not as an outcome of perception and thus part of the *explanandum*, but as part of the *explanans*.

The approaches introduced in this subsection are, in many ways, a methodological inspiration for the investigation of the problem of experienced simultaneity and its link to sensory delays introduced in section 8.3 that is presented in the following chapters 9-12.

8.2.5 Time Experience

In this subsection, I attempt to bring together the various findings presented so far. The objective of the previous summary is certainly not to give an exhaustive and cross-disciplinary account of time cognition. Each of the subsections presented introduces only a small number of selected findings from very different areas to do with time perception. However, sketching the landscape of methods, perspectives and findings, it is possible to identify connections and make them explicit, indicating the directions in which to venture when addressing a problem within the area of time cognition and time experience.

We can distinguish three dimensions along which the approaches presented here can be characterised: Firstly, there are the three levels of time cognition identified by the phenomenologists. Whilst the philosophical approaches sketched in subsections 8.2.1 and 8.2.2 span these levels, the empirical findings are more or less confined to the realm of the descriptive-narrative level of time experience (subsection 8.2.3) or the immanent level of time experience (subsection 8.2.4) respectively.

Secondly, there is a methodological continuum, from a mere first person approach (subsection 8.2.1) to a conceptual-contemplative approach making links to the physical world (subsection 8.2.2) to approaches that use second person methods (subsection 8.2.3) and third person methods (subsection 8.2.4) either proportionally or exclusively.

Thirdly, there is the ideological dimension, reaching from radical computationalist approaches (e.g., Eagleman & Sejnowski, 2002) over intermediary positions (e.g., Libet, 2004; Gibson, 1982, original in 1966) to the radical constructivist/enactive perspective (e.g., Varela, 1996).

Furthermore, it has been possible to see the issues mentioned at the beginning of this section, i.e., level of time experience, the intertwinement of time and space and the role of the known, the unknown and the possible (e.g., in the Aymaran culture or in visual prediction), to recur across all the accounts given. Having clarified that my personal convictions are on the constructivist end of the ideological spectrum, it is now possible to ask the following kinds of questions: what is the relation between pastness and knowledge? What are the appropriate methods to investigate experienced simultaneity? How do the findings thus obtained fit within the landscape drawn? What are the structural similarities and the relation between the findings on narrative-descriptive time and immanent time? What is the origin of experienced order and what do disruptions of this experience teach us?

After presenting the results from the combined experimental and modelling project on adaptation to sensory delays, some preliminary ideas on the nature of some aspects of temporal experience in the light of the presented results are given in chapter 12. At this point, the discussion of time perception and time cognition as a general topic, however, comes to a stop. The following section presents the problem of sensory delays and perceived simultaneity as a particular

issue in the study of time perception and outlines the hypothesis tested in this second part of the dissertation.

8.3 Adaptation to Sensory Delays and the Experience of Simultaneity

In a recent study, Cunningham et al. (Cunningham, Billock, & Tsou, 2001a) report patterns of adaptation to artificially prolonged sensory delays in human participants in a visuomotor task that are strikingly similar to those obtained in experiments with spatial displacement. Firstly, over training, the initially impaired performance is recovered and the annoying delay disappears from conscious experience. Secondly, re-adaptation to the normal condition is marked by a strong negative after-effect, i.e., participants' performance on the unperturbed condition without delay is worse after training with a 200 ms visual delay. Although their study focused on the behavioural aspects of the task, the authors report as anecdotal evidence that several subjects spontaneously reported that "when the delay was removed, the plane appeared to move before the mouse did - effect appeared to come before the cause" (Cunningham et al., 2001a, p. 533).

In the light of the previously presented insights, the results do not appear surprising. The reported research shows that our experience of nowness is constructed according to sensorimotor invariances and that the tacit flow of this experience can be transformed and brought to break-down or logical conflict. In particular, Libet's (2004) findings about neuro-behavioural latencies that are not part of our temporal experience make it appear plausible that systematic sensory delays will be integrated into perception of the present.

What makes these findings so interesting is that a similar adaptation effect had been hypothesised at several occasions, but had failed to occur in previous studies, which led Smith and Smith (1962) to conclude that adaptation to sensory delays is impossible in principle. They base their judgment not only on their own findings from experiments in which they imposed a visual delay on a drawing task (following the outline of a star), but also on reviewed related work by other researchers. Following up on Cunningham et al.'s reported results, Stetson et al. (Stetson, Cui, Montague, & Eagleman, 2006) tried to reproduce the effect in a minimalist psychophysics set-up but only produced partial readjustment of perceived simultaneity. A similar finding of partial stretching of crossmodal simultaneity perception in audiovisual perception was reported in (Fujisaki, Shimojo, Kashino, & Nishida, 2004). Other work that failed to produce adaptation to sensory delays includes experiments on telesurgery (Thompson, Ottensmeyer, & Sheridan, 1999), remote manipulation (Ferell, 1965) and a reported disruption of adaptation to spatial displacement by visual delays (Held, Efstathiou, & Greene, 1966). These results include studies on delays within the range of less than 100 ms to over 1s, from different modalities, from active and passive conditions in different behavioural domains. What is it about Cunningham et al.'s study that makes them different from those previous and later studies that failed to produce the described adaptation effect?

Cunningham et al. themselves hypothesise that the adaptation effect their study produced is due to the time pressure in the task that makes the delay meaningful for the solution of the task.

> "[I]t has been clearly demonstrated that sensorimotor adaptation requires subjects to be exposed to the consequences of the discrepancy [...]. Thus, it is of central importance to note that subjects in previous studies slowed down when the delay was

> present. [...] This is crucial because slowing down can effectively eliminate the consequences of the delay" (Cunningham et al., 2001a, p. 534).

This observation relies on a definition of adaptation that the authors adopted from Welch (1978) as *'semi-permanent'* change in perception that eliminates behavioural errors and/or the registration of a perturbation. Furthermore, the authors measure adaptation through the *negative after-effect*, i.e., the reduced ability to accurately perform the task without the perturbation originally induced as the "most common measure of adaptation" (Cunningham et al., 2001a, p. 533). Slowing down, as a compensatory strategy, may help to improve performance on a given task with sensory delays to a certain extent. It is, however, not a strategy that produces a negative after-effect or semi-permanent adaptation, but much rather a cognitive-logical compensation strategy. In a follow-up study in a multimodal task (Cunningham, Chatziastros, von der Heyde, & Bülthoff, 2001b), the reported adaptation could be reproduced under time pressure. In a delayed vestibular feedback condition (Cunningham, Kreher, von der Heyde, & Bülthoff, 2001c), however, only partial adaptation was found.

The findings on adaptation to sensory delays and its effect on experienced simultaneity form the starting point of the interdisciplinary project presented over the following chapters. I also adopted the hypothesis that time pressure is necessary to make the delay a meaningful discrepancy and induce adaptation. However, rather than testing the produced effect in ever more complex settings, I wanted to use the minimalist experimental and modelling approach described and developed in chapter 3 to find the minimal conditions for semi-permanent adaptation to sensory delays and distinguish them from one or several even simpler conditions in which the adaptation is not produced, in order to be sure that the conditions identified are really the minimal conditions. The project was conducted during the five months research stay at the GSP in Compiègne, using their experimental facilities (audiotactile feedback platform Tactos (Gapenne, Rovira, Ali Ammar, & Lenay, 2003)) and drawing on their experience with this kind of sensorimotor experiments. The model and initial results are described in the following chapter. Unfortunately, the initial hypothesis, i.e., that time pressure is the only factor necessary to lead to semi-permanent adaptation, was ultimately not confirmed. A close analysis of the data based on a simulation model of the experiment (chapters 10 and 11), however, led to new insights and hypotheses for further experimentation that are presented in chapter 12.

Locating this project along the dimensions identified in the previous section 8.2.5, this effect occurs at the level of the immanent flow of time at the scale of temporal object events. However, in contrast to Varela's neurophenomenological approach, my work investigates the quantitative properties of the experience of nowness and its sensorimotor correlates, not their distinction in terms of qualitatively different neurophysiological processes, even if both questions strongly relate. The experiment about simultaneity construction and sensory delays can be seen as a contribution towards establishing the exact scopes and limits of the immanent time scale and the plasticity of temporal experience within it. It is worthwhile pointing out that both the visual delay of 200 ms used by Cunningham et al. and the tactile delay of 250 ms used in my experiment are at the intersection of the time scales Varela (1999) identifies to characterise the primitive continuous flow of sensations and the immanent flow of discrete moments of presentness. Therefore, I believe, the visual delay is perceptible yet can be integrated into the experience of presentness. Evidence for

this hypothesis that exact magnitude matters in this way comes from Cunningham et al.'s (Cunningham et al., 2001b) experiment on adaptation to delays in a driving simulator. They compared adaptation to a 130ms, a 230ms and a 430ms delay and report a clear reproduction of the effect observed in (2001a) only in the condition with a 230ms delay, suggesting that the 130ms delay is too small to be registered and the 430ms delay is too long to be integrated. Further experiments would be necessary to test this hypothesis.

Concerning the methodological dimension, my project clearly focuses on third person methods, i.e., the measurement and scientific analysis of sensorimotor behaviour and performance on the task in which measurable quantities, such as negative after-effects are seen as indicators of the perceptual world and its adaptation. This approach is in line with Cunningham et al.'s (2001a), who report the interesting experiential phenomena observed as anecdotal evidence but concentrate on the measured behavioural success. As already mentioned in chapter 3, I had intended to also investigate the experiential dimension but have failed to engage sufficiently with the methodological difficulties and possibilities of such an undertaking.

As concerns the ideological dimension, this approach and the hypothesis adopted very much reflect the constructivist perspective underlying my research. A representationalist interpretation of the effect is exemplified in Stetson et al.'s hypothesis that "sensory events appearing at a consistent delay after motor actions are interpreted as consequences of those actions, and the brain recalibrates timing judgments to make them consistent with a prior expectation that sensory feedback will follow motor actions without delay" (Stetson et al., 2006, p. 651). This hypothesis and the consequent experimental set-up do not mention the significance of the delay or the nature of the task, they presume that adaptation proceeds automatic and based on syntactic-statistic properties (The authors do not make mention of the failure of the adaptation effect to occur in previous studies). Similary, the failure to produce the effect is not explained with respect to the sensorimotor dynamics or other meaningful factors, but it is hypothesised that "it may be that motor-sensory timing shifts of 100 ms are beyond the hardware limitations of the calibration mechanisms" (Stetson et al., 2006, p. 656). This example should illustrate in how far Cunningham et al's hypothesis, which I adopted for my work, is of a fundamentally different nature: it relies on the significance of the perturbation and characterises these in terms of closed loop sensorimotor dynamics.

A last issue to mention in this section is the role of identity in the investigated phenomenon. The disruption of temporal experience spontaneously reported by Cunningham et al.'s (2001a) subjects is surprising to them because aspects of one meaningful action are temporally torn apart such that what is experienced as the effect (i.e., movement of the plane) precedes what is experienced as the cause (movement of the mouse). Therefore, the experiential effect is not actually the distortion of temporal order of two otherwise simultaneous temporal object-events (different in space but identical in time), but much the separation of crossmodal aspects of an experience that are usually experienced as one identical object-event (in both time and space). While experiencing the delay of an effect is something we are used to from everyday life, the advance of effect appears to us as a logical contradiction. This distinction (i.e., crossmodal identity vs. simultaneity) is an issue that will have to be born in mind in future experiments.

Chapter 9

An Experiment on Adaptation to Tactile Delays

The experiment on human adaptation to sensory delays was conducted in collaboration with the Groupe Suppléance Perceptive (GSP) during a five months research stay at the Technological University of Compiègne. In particular, Olivier Gapenne, Dominique Aubert, John Stewart and Charles Lenay have helped me crucially to develop and conduct the experiment. The objective of the experimental study was to reproduce the findings reported in (Cunningham et al., 2001a, cf. chapter 8) in a minimal sensorimotor task. The experiment should test the hypothesis that time pressure in the task is necessary in order to produce semi-permanent adaptation to sensory delays as defined in the previous chapter (section 8.3), which is characterised by a negative after-effect when returning to the original pre-adaptation condition. Furthermore, I recorded the participants' motion and sensation in order to be able to present data about the behavioural and sensorimotor dynamics of adaptation.

This chapter outlines the experimental protocol and describes the design decisions made (section 9.1). It then presents the results of the experiment (section 9.2), which seem disappointing at first glance (no statistically significant adaptation in terms of performance) and briefly discusses them as they stand alone. After presenting the ER simulation of the experiment and the findings it generates in chapter 10, the experimental results are revisited in chapter 11, where a further analysis of the recorded data, informed by the synthetic results from the simulation model, is performed and the experiment is re-evaluated. In chapter 12, I draw the conclusions for this second part of my DPhil as concerns the question of adaptation to delays and perceived simultaneity. This presentation broadly follows the order in which I worked on the different parts.

9.1 Experimental Set-Up

My experiment was implemented using the tactile feedback platform Tactos (Gapenne et al., 2003) that the GSP have developed. It links participants' motion in a simulated environment (movement of mouse, stylus, etc.) systematically to patterns of tactile stimulation on a four-by-four Braille display (see figure 9.1). The Tactos system can be used as a perceptual supplementation device as outlined in chapter 3 and (Lenay et al., 2003), to investigate the perceptive qualities of previously unfamiliar sensorimotor couplings. The advantage of using a simulated environment for this kind

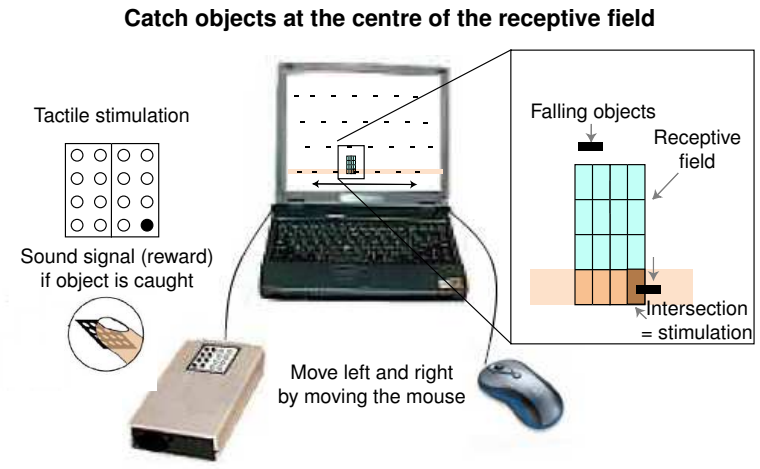

Figure 9.1: The Tactos tactile feedback platform. Task: Objects have to be located in the centre of the perceptive field when they reach the bottom line.

of perceptual supplementation experiment is that it is very easy to modify, control and record the parameters of artificial sensorimotor couplings. For the purposes of my research, it is also advantageous that the simple virtual environments used in the experiment serve as environments for the artificial evolution of robotic agents, as outlined in chapter 3.

The aim was to find the minimal conditions under which the semi-permanent adaptation to sensory delays takes place that Cunningham et al. (2001a) report and to distinguish them from nearly identical experimental conditions in which this adaptation does not occur, in order to be able to say something about the minimal conditions for this adaptation to take place. In the experiment presented in this chapter, I have, however, not achieved to reproduce Cunningham et al.'s findings in the simplified set-up here presented, i.e., the main hypothesis is not supported by the data collected.

The experimental set-up in Cunningham et al.'s experiment is already rather minimal: Participants move along one dimension (mouse movement to the left and right) in order to avoid evenly spaced obstacles. These obstacles are arranged in a field that participants traverse at a fixed linear velocity from the bottom to the top (i.e., orthogonal to the direction in which they can move with the mouse). However, even if the logic of the task is not very complex and the possibilities to act are rather restricted, the sensory inputs in this task are rather rich. Besides the non-delayed proprioceptive/reafferent feedback about self-movement and the position of the mouse, the screen provides a visual representation of the field of obstacles and the location of the airplane. The airplane is delayed by an additional 200 ms in the delay condition to which participants are supposed to adapt. The visual sense is a very complex sense and it is difficult to explain what in the complex and informationally rich representation has been exploited to solve this task. Therefore, in order to find the minimal conditions for adaptation to sensory delays and be able to analyse the sensorimotor dynamics of adaptation, the most important part was to simplify the sensory component of the task by transferring it to the tactile domain (and blindfolding subjects). In the following, I describe different aspects of experimental set-up used and explain certain design choices.

9.1.1 Task

In Cunningham et al.'s experiment (2001a), the movement along the vertical dimension is not under the participants' control (fixed linear velocity), whilst movement along the horizontal dimension is entirely under the participants' control (no externally induced motion of plane or obstacles along the horizontal dimension). We adopted the same task organisation for several reasons. Firstly, this organisation delimits degrees of freedom and hence makes the sensorimotor behaviour tractable and mathematically more manageable. Secondly, these distinct sensorimotor couplings can be linked to the distinction between time and space through *reversibility* observed by Kant (cf. chapter 8) and thereby eases conceptual analysis and interpretation of behavioural strategies, making the task and its interpretation in the context of the debate opened up in the previous chapter more semantically manageable. Thirdly, a similar organisation is adopted in the catching and avoidance tasks Beer frequently adopts for his evolutionary robotics simulation studies of minimally cognitive behaviour (e.g., Beer, 2003, 1996). This similarity allows us to compare the results from the simulation of the experiment with those of earlier ER work and possibly even makes a transfer of some of Beer's elaborate methods of dynamical analysis possible.

In order to simplify the sensory dimension of the task, it was transferred to an audio-tactile environment. Participants were blindfolded and received tactile stimulation via a Braille display (see exact specification below) and auditory signals indicating object velocity and reward for successful behaviour. As already stated, we adopted Cunningham et al.'s (2001a) conjecture that time-pressure is essential for adaptation to sensory delays to take place. Obstacle avoidance in a dynamic environment, as employed in Cunningham et al.'s study, is one of the simplest possible tasks that require real-time interaction. Initially, we had planned to use an obstacle avoidance task as well. However, participants in piloting experiments became extremely distressed when tested in the obstacle avoidance task (to the point that they felt unable to complete the experiment). It seemed that the combination of just receiving negative feedback, to indicate collision, and the unfamiliar sensorimotor couplings were at the root of the distress participants suffered. They felt punished, without feeling they could do anything about it. Therefore, we changed the task to a catching task. Catching objects is an equally simple task that requires real-time interaction, but it is a much more positive task, because reward upon catching objects is indicated, rather than punishment upon collision.

The simulated environment contains objects of size 1×4 units that are evenly spaced (28 distance units) in the horizontal dimension (see figure 9.2 (A); what these distances mean in human arm movement space is specified below). The space is infinite and rows of objects fall down at one of seven different fixed velocities ($v_o \in \{0.004, 0.006, 0.008, 0.010, 0.012, 0.014, 0.016\} units/s$) from a distance of 25 space units above the bottom line. The absolute spatial location of the objects to catch follows a sequence of four positions that is repeated (see figure 9.2 (B)), a technical choice which, given the Tactos system, makes implementation much easier. The *perceptive field* of the participants is of size 16×8 and is subdivided into 16 areas of size 4×2 that correspond to the 16 pins on the Braille display, i.e., intersection of one of the perceptive areas with one of the falling objects activates the corresponding pin on the Braille display.

The structure of the environment thus implies that at any point in time a maximum of three neighbouring pins is activated on the Braille display. All 16 pins of the Braille display are used in

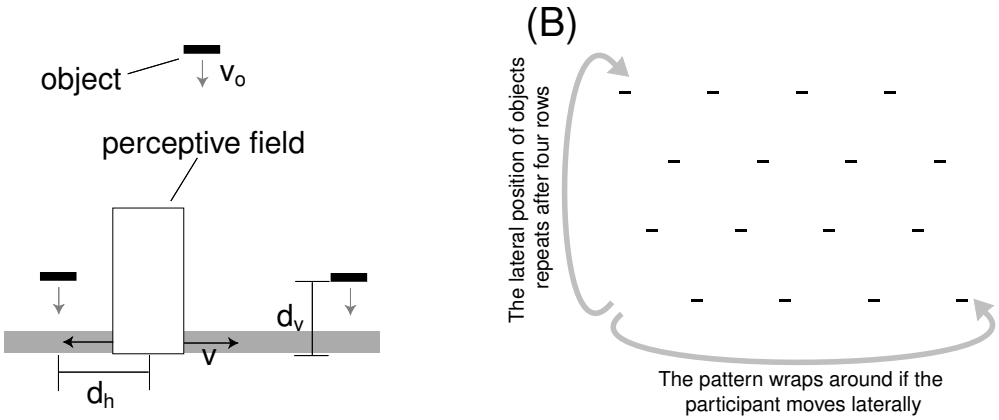

Figure 9.2: (A) Illustration of the simulated experimental environment (B) The repetitive lateral displacement of rows of objects.

the task. The four rows (temporal dimension) indicate the horizontal distance to the object which corresponds to the time available to position oneself below the object. The four columns (spatial dimension) indicate the horizontal distance to the object that is determined through participants mouse movements. Is this really the *minimal* sensory complexity possible? Previous work by the GSP has established that, to a degree, enhanced motion possibilities can compensate for receptive field parallelism (e.g., Lenay, 2003). It has to be born in mind that such a active perceptual strategy of distances and velocities would possibly complicate the sensorimotor strategy for performing the catching task. Introduction of a sensory delay, consequently, could interfere in more complex and less understandable ways with successful performance in the task. Therefore, including structural information on the participants' *motion* and the *velocity* of falling objects into the patterns of stimulation appeared an appropriate design decision.

The exact definition of the task is to move along the horizontal dimension such that the object is approximately at the centre of the receptive field (distance between the object margin and the receptive field centre < 4 distance units) when it touches the bottom line. This implies that in each row the participants can catch at most one object. If this task is accomplished, an auditory reward signal is triggered to indicate success (an additional bit of information). In piloting studies, we observed that some participants are not very accurate in discriminating the exact patterns of activation, even if presence/absence of stimulation is easily perceived. Such participants sometimes fail to locate themselves directly beneath the object and get frustrated because they cannot tell whether the object is in the centre of the receptive field or at the margin. I decided to keep up the requirement of centring the object nonetheless, for several reasons. Firstly, participants concentration was increased this way. Secondly, and more importantly, counting a touch at the margin gave way to sloppy behaviour. The decrease in spatial accuracy of the task made the introduction of a delay a less drastic perturbation. On the other hand, simply slimming the perceptive field in order to avoid confusing stimulation at the perceptive field margins turned out to make the task too difficult, participants struggled to locate the objects in the first place.

Taking into consideration the velocities at which the objects fall, the described structure of the task implies that there is a time window to accomplish the task between the moment an object first

becomes perceptible to when it reaches the bottom line of between exactly 1*s* for the fastest objects and 4*s* for the slowest objects in the condition without delay and 250 ms shorter for the condition with delay respectively. During the phases of performance measurement, only the five fastest velocities have been used. These velocities have been identified during piloting to be sufficiently fast to induce a time pressure in the task, but not too fast to accomplish it (see details in section 9.1.3 part below). The vertical motion of falling objects is paused for 100 ms in between each presentation of a row of falling objects. This gives participants the possibility to gather and prepare a little bit for the next row of objects, while not impacting dramatically on the time necessary to conduct the study.

When changing the task from an avoid task to a catch task, one difficulty encountered was that subjects sometimes do not realise that objects are falling. In an avoid task, the objects seek the participants, stimulate them and force them to react. In a catch task, however, stimulation has to be sought and a lack of stimulation can indicate either absence or distance of objects. In order to mitigate this problem, an additional auditory signal was introduced, a sound emitted every four distance units an object falls in order to indicate the presence and velocity, even if objects are not currently in the perceptive field. In terms of informational content, this auditory signal repeated information potentially available via the Braille display, i.e., a sound indicated a potential change in row on the Braille display. Hearing this signal, participants were encouraged to search for objects to catch which they knew were passing on the side. As mentioned before, participants were *blindfolded* to avoid distraction by visual stimuli and to control sensory inputs.

Movement on the horizontal axis was recorded using an optical mouse. The vertical component of mouse movement was simply ignored. Due to a miscommunication, the operating system's (MS windows XP) acceleration curve (cf. Microsoft Corporation, 2002) was applied to the mouse movement, which is not usually meant to be the case in the Tactos system. This impacts on the data analysis, but not on the general logic of the task. The relevant parts of the analysis have been conducted as if the mouse movement recorded was absolute position (before detecting the misunderstanding). These effects of acceleration pre-processing are pointed out where relevant in the presentation of the data in this chapter and chapter 11.

Even though, with the acceleration curve applied, the correspondence between desk space covered and movement in the simulated environment is ambiguous, moving the mouse with normal velocities, the approximate relation between mouse movement and receptive field movement is 0.42 cm corresponding to the 28 units that separate objects within one row. In comparison, using the mouse in the same configuration to navigate as normal on the screen, ≈ 18 objects would be crossed traversing the screen from left to right, which means that the mouse is rather fast in the simulated environment.

Both the choice of a fast mouse and of an optical mouse introduce a decrease in spatial accuracy into the task and spatial accuracy very important for the behaviour observed (see chapter 10 and 11). These choices are, however, not fully without justification. An optical mouse is preferable to a ball mouse, because a ball mouse can get blocked and there is no way for participants to detect this additional perturbation; the experimenter would have to interfere. The alternative of using a graphics tablet that tracks a mouse or a stylus very exactly may in the end have been a preferable choice. However, it seemed problematic at the time the paradigm was developed be-

cause blindfolded subjects would have difficulties realising when they exceed the graphics tablet, given that the virtual space is infinite, or, running up against a limit, they may pick up the mouse and put it at the other end, unjustly presupposing that such a movement would not be recorded as is the case for the more common ball or optic mice.

Choosing a very fast mouse was due to a different observation made during piloting: As soon as participants wear a blindfold, the amplitude of their movements decreases drastically without them noticing. Even when pointing out this behaviour to the participants, they remained in disbelief and insistently made very small movements. With a slower mouse, therefore, many subjects move within a very small area of the space, whilst being convinced that their scanning movements cover the entire screen. By making the mouse fast, in contrast, subjects scanned at a magnitude large enough to actually localise an objects and make contact with it. At the same time, a mouse that is very sensitive makes small spatial adjustments very difficult, which increases the spatial inaccuracy of the set-up chosen.

9.1.2 Delay

In the experiment, both the tactile and the auditory signal were delayed by 250 ms additional to the inevitable delay of $\approx$ 35 ms the computer induced and that is present (though not perceivable) in all conditions. Again, this choice is due to several considerations.

Cunningham et al. chose a visual delay of 200 ms in their visual delay task in (Cunningham et al., 2001a). In a different multimodal driving simulation task, they worked with different delays of 130 ms, 230 ms and 430 ms and found the negative after-effect to be only clearly produced in the 230 ms group (Cunningham et al., 2001b), suggesting that delays of $\approx$ 200+ms are - intermodally - the kind of magnitude to delays subjects can adapt to semi-permanently.

A different issue in choosing the delay length was the experience of the delay, as I was interested in this dimension as well. In piloting experiments, differences in subjective sensitivity to delay magnitude were detected. Some subjects did not have the experience of delayed sensation even if delays were very long (e.g., 500 ms), because they could not fully *make sense* of their actions and the results they produced, given the complete novelty of the sensorimotor coupling. With a tactile/auditory delay of 250 ms, however, most participants report that they perceive the sensor feedback as delayed (or at least report that they register a perturbation, if not prompted to conceptualise it as delay), whereas smaller delays led to more ambiguous experiences.

Finally, there are technical reasons for choosing a delay that is of a considerable magnitude. Firstly, the system has inaccuracies of $\approx \pm$ 15ms in processing between the systems involved (simulation, mouse, Braille display, sound). Secondly, even though the system is programmed to write the sensorimotor data to log files every 15 ms, it sometimes skips one or two such 15ms time steps in favour of maintaining task coherence in time. Therefore, in order to be able to say something meaningful about the sensorimotor behaviour with and without delays, it seemed necessary to take a value significantly larger than the 30ms inaccuracy in processing or the 45ms inaccuracy in writing. 250ms hence was the value chosen to be small enough to allow adaptation, yet be large enough to be perceptible and meaningful despite inaccuracies.

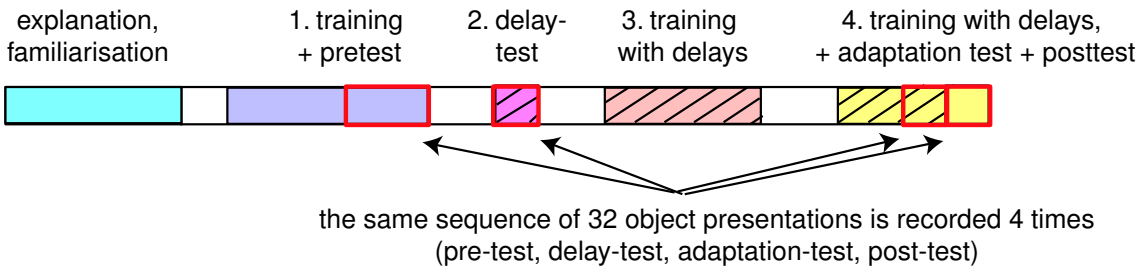

Figure 9.3: Visualisation of the experimental protocol. The five experimental phases in different colours. Plain: no delay; striped: delay. Red box: recording of fixed sequence of 32 objects with variable velocity (longer in pre-test because it contains the overlapping sequences for the three velocity groups).

9.1.3 Protocol

The experiment consisted of five experimental phases that, altogether, lasted 30-45 minutes (see figure 9.3).

In the first phase, participants were familiarised with the system and the task. They were made to catch objects at the slowest velocity until they caught them reliably, which would take between 2 and 10 minutes. Data from the familiarisation phase was not recorded.

The *first experimental phase* lasts 6:40 minutes. Participants perform the task being given objects that fall at increasing velocities (first four minutes) before the task goes over to a sequence of variable velocity objects. This experimental phase serves three purposes: a) to further train subjects on the task with different velocities b) to record performance to rate participants and assign them to one of three groups of participants c) to measure performance on a fixed sequence of object velocities at the end of this phase to record the base-line performance without delayed sensory feedback (pre-test).

Regarding the classification of subjects, I had specified a formal criterion that if participants caught at least 80% of the objects of the second fastest velocity, subjects were assigned to the group 1 (fast), if they caught at least 80% the third fastest velocity, they were assigned to group 2 (average) and all the remaining participants were assigned to group 3 (slow). In practice, however, I sometimes deviated from this formal criterion, if poor performance was either clearly due to external factors or to misunderstandings. Given the noisy environment in which the study was conducted (see remark below), there were some cases of external disruptions, which is not ideal. Out of the 20 subjects I recorded, 3 were assigned to group 1, 9 to group 2 and 8 to group 3.

In order to compare the performance at different stages of the experiment, I used three different fixed sequences of 32 objects for the different groups. Group 1 was only tested on the top three velocities, group 2 was not tested on the fastest velocity, but on velocities 2-4 and group 3 was tested on velocities 3-5, omitting the top two velocities. All three sequences were passed at the end of the first experimental phase in order to have the pre-test data for all participants already measured irrespective of the group they were later assigned to.

The second experimental phase was a measurement of the perturbation that the introduction of the delay meant. Participants were given the same sequence of objects they were measured on during the pre-test (according to their group), but the tactile and auditory feedback was delayed

by an extra 250ms. This phase, which was also recorded, lasted only ≈ 1:30 min (1:15 for group 1, 1:25 for group 2 and 1:40 for group 3) and is referred to as 'delay-test'.

The third experimental phase served to train subjects to perform the task with the delay. This training was not always sufficient, training continued in the fourth phase, but more than six minutes of performing the task are perceived very straining by participants, so I had to introduce a break to counter fatigue. During ≈ 5 minutes (5:10 for group 1, 5:05 for group 2, 5:15 for group 3), participants were presented with objects at increasing velocity up to the top velocity for their group. Sensorimotor data from this phase was recorded.

In the fourth and final experimental phase, participants were given a further chance to adapt. They were again given increasing object velocities, even if velocities increased more rapidly. After reaching the top velocity, they were tested on the three top velocities for their group in random order. At the end of this random velocity sequence, participants were again presented with the sequence of 32 object velocities they had first been measured on to be able to quantify in how far their performance improved (adaptation-test). Then participants heard a sound signal (a 'gong'), the simulation paused for a second, and they were presented with the same sequence of object velocities, but without the sensory delay. Participants had been informed in advance that the 'gong' and the break mean that the task returns to the no delay condition they had been performing on originally. This condition during which behaviour is identical to behaviour in the pre-test is referred to as the post-test, in which a negative after-effect is predicted. Altogether, this fourth experimental phase lasted ≈ 5 minutes (4:30 for group 1, 5:05 for group 2 and 5:30 for group 3)

Participants were told in advance that the perturbation they suffered was a delay, even if some of them spontaneously reported that they did not perceive it as a delay, just as 'something wrong', or that they only realised it was really a delay when they returned to the original condition.

By virtue of this sequence of experimental phases (figure 9.3), every participant was measured on the same sequence of objects four times, once after having gotten used to the task but before training with delay (pre-test), once before training with delay but with delay (delay-test), once after training with delay (adaptation-test) and once after training without delay (post-test). 32 objects, which corresponds to ≈ 1:30 minutes of performing the task (see times in delay-test), is a time frame short enough to register the only temporary negative after-effect and, at the same time, long enough to allow speaking meaningfully about performance average. It is important to realise that, even though the velocity sequence used varies between groups of participants, each participant's adaptation was tested on exactly the same condition. Some analyses of the effect of this classification are discussed in chapter 11.

Performance F (in allegory to 'fitness' in ER simulations) is thus defined as

$$F = \frac{1}{32}\sum_{1}^{32}|d_h| < 4 \tag{9.1}$$

where d_h the distance between perceptive field centre and object margin at the time the object reaches the bottom line.

The comparison of the performance during these four presentations of the fixed sequence of 32 objects during different phases of the experiment is most important for the experiment. It is at the heart of the data analysis in the following section 9.2, as well as for the further data analysis in chapter 11. The *main hypothesis* in this experiment was that the described set-up would produce a

negative after-effect, i.e., that performance deteriorates between the pre-test and the post-test, even if the conditions are identical. Furthermore, it was expected that the introduction of the delay perturbs the performance (comparing pre-test and delay-test) and that the training with delays would lead to a (partial) improvement of performance (comparing delay-test and adaptation-test). However, these two latter effects are in themselves not sufficient to give evidence for semi-permanent adaptation, even though they may be considered necessary.

A different, more diffuse aspect investigated was that the changes in performance and/or experience are associated with measurable changes in the sensorimotor behaviour or modification of strategy for solving the task. One effect that was to be expected is that upon the introduction of a delay, there would be an *overshooting* of the object location, whilst, after adaptation, when the delay is removed, this effect would be inverted, as a negative after-effect similar to those observed for spatial displacement (see (Bedford, 1993; Welch, 1978) for a review). However, these effects were not fully phrased out as hypotheses but proceeded in a more exploratory way. As argued in chapter 3 section 3.4, this post-hoc data analysis focuses on descriptive aspects of the data, rather than explanatory. New theoretical insights gained from this kind analysis may require further experiments or additional control conditions.

The experiment was performed on 20 unpaid subjects of different age-groups (mostly graduate students) and both sexes. Our experimental protocol (Rohde & Gapenne, 2006) had won the experimental protocol competition of the French Cognitive Science conference (ARCo'06), the prize of which was to be allowed to run the experiment on the participants of the conference. Even though this prize meant that the experiment could be run on a large number of participants in a comparably short time span, it was not unproblematic. The space set up was rather cold and not properly partitioned from the main university entrance hall, such that some participants were occasionally perturbed in their performance by spectators or noises from the environment.

An additional difficulty was that the first three subjects recorded suffered a spatial irregularity (increased inter-object spaces) that was due to the transfer of the system from the machine it was developed on to a portable computer and that was corrected for as soon as it was detected. The irregularity seemed sufficiently small for me to use the data from these three recordings in the overall evaluation of the experiment. In the analysis of the sensorimotor behaviour, however (e.g., mean velocity etc.), the data from these three trials has been neglected, because it cannot be assumed to follow the same regularities.

The reported issues about the sub-optimal conditions under which the experiment was conducted and the unintended inclusion of the acceleration curve are clearly not ideal. It has to be born in mind, though, that the current dissertation is, primarily, a methodological dissertation that focuses on how simulation modelling can be tied into an experimental and experiential framework. Even if further experiments may be necessary to make strong and scientific claims about the sensorimotor basis of perceived simultaneity, the results presented in this second part of the dissertation serve perfectly well to illustrate and demonstrate that the scheme proposed is useful and promising.

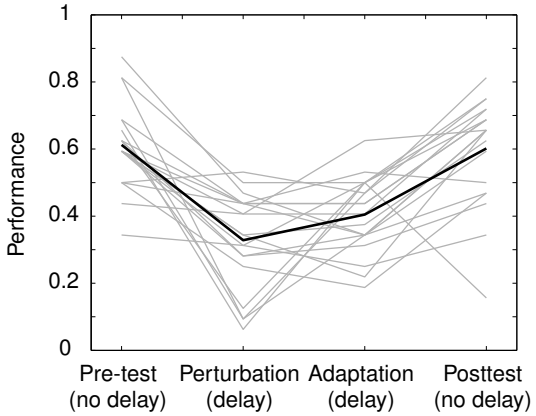

Figure 9.4: Participants' performance during the different phases of the experiment (Pre-test, delay-test, adaptation-test, post-test). Black line: Average performance. Grey lines: Individual Performances.

9.1.4 Questionnaire

I had developed a questionnaire for subjects to evaluate the experience of the delay and simultaneity along different dimensions after each of the experimental phases. This questionnaire, however, had already been discarded during the piloting phase of the experiment, because subjects had difficulties accessing and describing their experience.[1]

As already argued in chapter 3 section 3.5, naïve humans are generally only to a limited extent able to describe their experiences. A worthy description requires, e.g., extensive interviews by a trained and experienced interviewer familiar with techniques of consciousness taking (e.g., Petitmengin, 2006; Vermersch, 1994). For the present purposes, this kind of approach may not even in future experiments be necessary or the most suitable. The kind of primitive change in experience that Cunningham et al. (2001a) report to have occurred and that lead to spontaneous report by several subjects (anecdotal evidence) appears to be better accessible through the measures associated with the discipline of psychophysics (cf. section 3.5).

9.2 Results

As outlined in the previous section, the main hypothesis was phrased in terms of how the performance varies on the repeated sequence of 32 objects that was presented at different phases of the experiment. Most importantly, we expected a decline in performance between pre-test and post-test (negative after-effect). Furthermore we expected that the introduction of a delay would lead to a decline in performance and that the training with delays would, at least partially, alleviate this decline. figure 9.4 depicts the performance profile of participants (individual and average) during the different phases of the experiment.

The introduction of the delay did clearly lead to a deterioration of performance (average performance drop of 0.28 comparing the pre-test with the delay-test $p = 0.7 \cdot 10^{-7}$). Performance markedly recovered with training (mean improvement of 0.08 comparing the delay-test with the

[1] The questionnaire featured questions such as 'On a scale from 1-5, how difficult do you find the task?' and 'Can you briefly describe how you solved the task?'.

Table 9.1: Average value across 20 participants for several variables that describe the trajectories.

	pre-test	adaptation-test	delay-test	post-test
1.) Performance	0.6125	0.3281	0.4047	0.6016
2.) Touched, but not caught (% of caught)	31.12	62.86	52.90	31.17
3.) Average distance at end of trial	1.5092	3.7555	2.6746	1.6232
4.) Average velocity	22.5436	23.2944	17.9297	20.0883
5.) Proportion of time in motion	0.3627	0.3640	0.3631	0.3620
6.) Average velocity while moving	62.6843	60.5646	48.8942	56.0315
7.) Average time spent before stopping	0.5455	0.4831	0.4847	0.4669
8.) Ratio of objects caught with $v = 0$	0.8594	0.5953	0.6875	0.8438
9.) Number of crossings of the object	1.3703	1.3875	1.2469	1.2453

adaptation-test), though this recovery was not statistically significant ($p = 0.067$). At these earlier phases of the experiment, some individuals followed already very unexpected patterns in their performance: Some maintained their level of performance when the delay was introduced, or even got slightly better with it, whilst others got worse with training (see figure 9.4). It seems to be this large inter-individual variance and atypical individual trajectories that account for the fact that improvement with training (delay-test vs. adaptation-test) is not statistically significant, rather than a general trend not to adapt to the induced perturbation.

For a negative after-effect, on the other hand, there is no evidence whatsoever. Performance decreased between pre- and post-test by a negligible 0.01, a change that is far from statistically significant ($p = 0.812$); the participants' performance stayed literally unaltered, which means that the main hypothesis has not been confirmed. However, rather than putting the head in the sand at this point, I started to analyse the behavioural trajectories more closely.

Table 9.1 gives the average value across individuals during each of the experimental phases for a number of the variables that have been investigated (for instance, the average velocity, proportion of time spent immobile, distance from object centre upon end of trial, ...). It is important to notice that all velocities given on this table are modulated by the mouse acceleration curve as mentioned in section 9.1.3. These velocities do not actually represent the average velocities corresponding to the movement of participants but the average over slightly weighted velocities that overestimate/underestimate the actual ones.

Many of the investigated variables followed a pattern of change that was similar to the one observed for the performance (1.), i.e., values were qualitatively identical between pre- and post-test (column 1 and 4), even if the introduction of the delay induced a perturbation (column 2) that was partially restored over training (column 3). For instance, variables 2.), 3.) and 8.) followed this pattern. However, others variables followed a different pattern, i.e., either a stepwise (with training) or a gradual (across all phases) change in one direction that differed between pre- and post-test, even if these differences were not statistically significant (variables 4.), 6.), 7.) and 9.) followed this pattern). These changes already suggested that a transformation may have happened across the training with sensory delays, an impression that is supported by the mere optical appearance of plotted trajectories. Figure 9.5 gives an example of a subject that exhibits a different

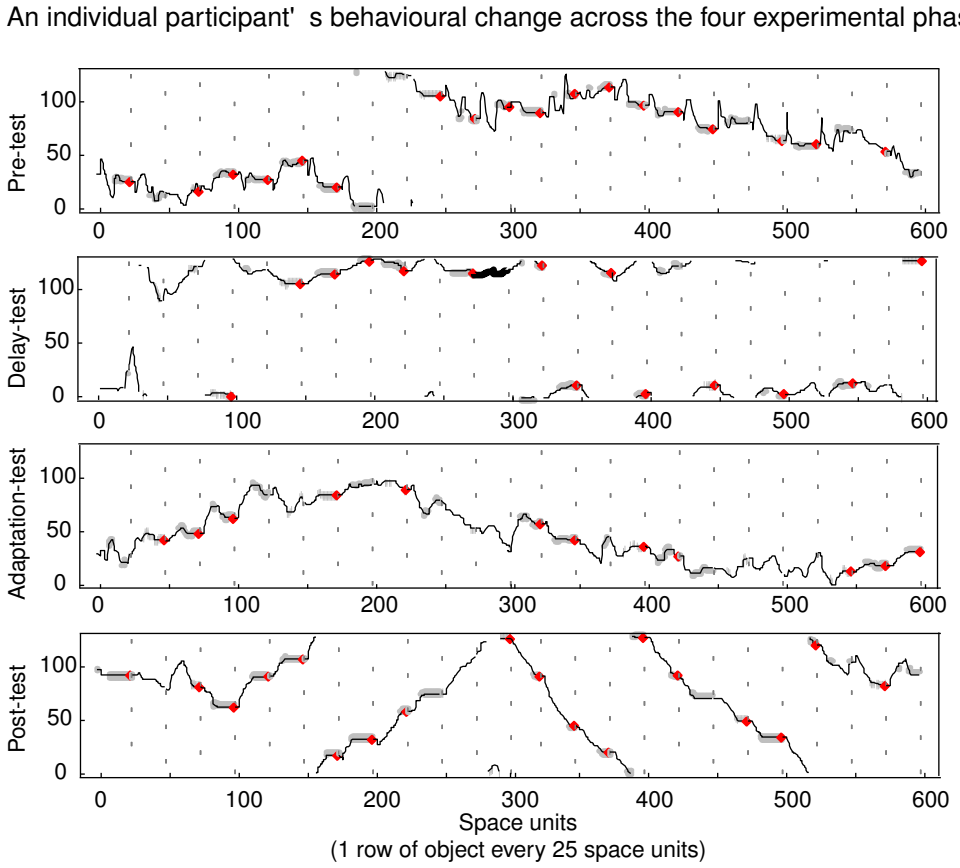

Figure 9.5: Trajectories (variable object velocity not represented) of an example participant over the course of the experiment. Even though the performance is identical in pre- and post test (75%), the behaviour has clearly been transformed over training. The slow movement and long halts during the post-test more resemble the behaviour when the delay was introduced, whilst the ongoing micro-swaying during the pre-test is somewhat restored in the adaptation-test, even if the performance is not.

behaviour in pre- and post-test, even if this difference is not reflected in performance.

There are two ways in which the refutation of the hypothesis could have been a trivial failure of the prior suppositions. Firstly, it could have been strictly impossible for participants to adapt to the delay. Secondly, an adaptation could have been a merely cognitive adjustment of strategy that is not semi-permanent and therefore does not produce a negative after-effect. The mentioned differences in behaviour between pre- and post-test suggest that neither of these two trivial explanations applies, because, in both cases, we would expect the behaviour to be the same in these otherwise identical conditions in all respects, not just in performance.

In what follows now, I report impressions and intuitions that were generated when I studied the data in depth, trying to understand it. These findings are not proper scientific findings in the sense that I would include them into a scientific publication. They are, however, important in telling the story of the interdisciplinary project as a methodological adventure. This formation of intuitions led over to the simulation model presented in the following chapter, which then generated the concepts and measures for the more tangible post-hoc data analysis presented in chapter 11.

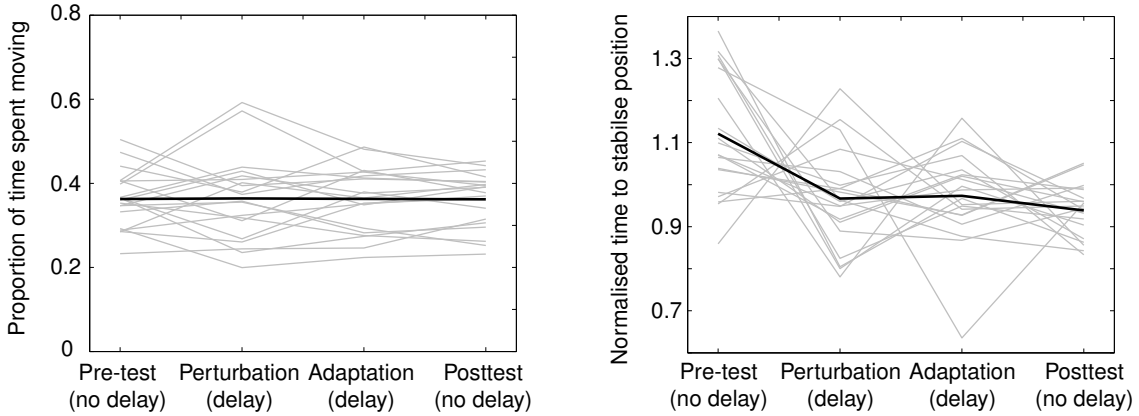

Figure 9.6: Participants' behaviour changes during the different phases of the experiment. (A) The average proportion of time spent moving (variable 5.) in table 9.1) hardly varies at all across the experiment. Looking at inter-individual differences, however, reveals that there are substantial changes which cancel each other out and that 9 out of 20 individuals show a negative after-effect in this variable in this sense. (B) The average time needed to stabilise and come to a stop (normalised by the average value per participant for this variable) decays significantly ($p = 0.1 \cdot 10^{-4}$) over the course of the experiment. Changes in velocity follow a similar pattern.

A closer look at inter-individual variation of behavioural/performance profiles leads to further questions about the nature of the results. Figure 9.6 plots the mean and the individual variation of the proportion spent in motion (A) and the normalised time needed to come to a halt (B). This variables correspond to variables 5.) and 7.) in table 9.1. Even though the average proportion of time spent moving changes hardly at all, figure 9.6 (A) shows that for most subjects, the introduction of the delay leads to a substantial change in this value which is then reverted over training - only that the perturbation leads to an increase in time spent moving in some participants, while in others it decreases it, such that the effects cancel out when calculating the average. For nine of the 20 participants, the change in proportion of time spent moving follows the zigzag profile hypothesised for the performance (i.e., if the delay leads to an increase in proportion spent moving, it will decrease over training and increase again in the post-test or the other way around). A similar tendency can be observed for the mean time to stabilise (see figure 9.6 (B); eight out of 20 participants followed a zigzag profile for this variable). Additionally, if this variable is normalised by the average value for each participant, there is an overall average decay which is significant ($p = 0.1 \cdot 10^{-4}$).

In conclusion it could be seen that there are inter-individual differences between participants as to if and how they adapt to the sensory delays introduced. If there is anything useful to be said about the data, it is necessary to identify those dimensions that mark general behavioural modification phenomena and distinguish them from others that are variably realised across participants. In the pursuit to find such meaningful variables and measures, I tried to classify the participants, initially on the basis of intuitions, in order to then see if this classification correlates, e.g., with the velocity groups or is immediately associated with a descriptive variable in the data. Simply on the basis of looking at visualised trajectories, occasionally consulting variables such as those listed in

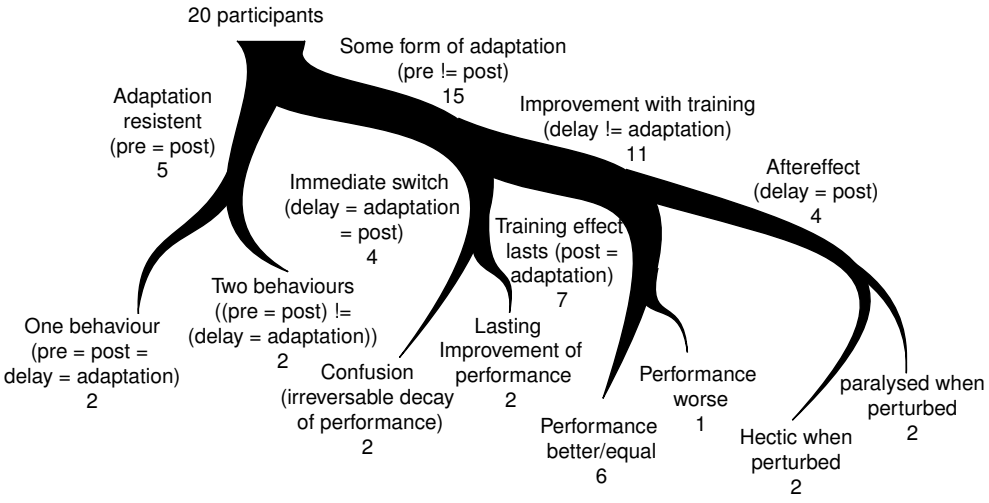

Figure 9.7: An intuitive classification of participants' behaviour based on visual representations of trajectories and the descriptive values in table 9.1. Performance profiles are very variable. For instance, four participants show no change from pre- to post-test. 11 participants undergo a change of behaviour during training with delays, but only four experienced a negative after-effect in the sense that the delay-test resembles the post-test.

table 9.1, I generated the decision-tree-like structure depicted in figure 9.7. The branches of this tree are distinguished by decisions such as 'Is the behaviour in the post-test at all different from the behaviour in the pre-test?' and 'If yes/if no, does the behaviour change at all when the delay is introduced/across adaptation?'. The most important result of this procedure is that 15 of the 20 participants were rated to change their behaviour between pre- and post-test and thus seemed, at least at first glance, to have undergone some semi-permanent behaviour transformation, even if the number of objects caught did not reveal it (such as the participant whose behaviour is depicted in 9.5). The exact nature and moment of the transition between pre-test and post-test behaviour, however, again seemed little homogenous (excessive sub-branching in the right hand branch of the decision tree in figure 9.7).

One of the main objectives of this project was to test the usefulness of combining and co-developing minimalist experiments in human sensorimotor adaptation and ER simulation modelling (as outlined in chapter 3 section 3.6), with the merits of ER simulation modelling being to clarify issues in sensorimotor dynamics that are too complicated to understand otherwise. At this moment, the data set to be analysed had become immense, particularly if dependencies between variables in different sub-groups of participants are taken into consideration. A simulation of the task to complement the experiment had been developed alongside the experiment. After 'getting a feel' for the data, it was a good moment to pause the data analysis and take a closer look at the simulation model of the task. By investigating the sensorimotor dynamics in idealised and unbiased settings, I wanted to understand the dynamics of behaviour modification in order to find variables to either ground the classification depicted in figure 9.7 or to find dynamical principles that hold across these sub-groups. The following chapter presents the model and its results, before a profound data analysis in the light of the simulation results is presented in chapter 11.

Chapter 10

Simulating the Experiments on Adaptation to Sensory Delays

This chapter describes the Evolutionary Robotics simulation of the experiment presented in the previous chapter. Even if the model had been developed alongside the experimental set-up, the major part of this work has been performed after the conduction of the experiment, with the objective to clarify the ambiguous data and the sensorimotor dynamics of adaptation and to generate hypotheses for further experiments. The results presented here have been in part published in (Rohde & Di Paolo, 2007).

The already gathered data is revisited for further post-hoc analysis in chapter 11 in order to test the predictions the model generates about sensorimotor principles associated with certain strategies. Apart from these concrete quantitative predictions, the model generates significant conceptual insights about the meaning of delays in different kinds of sensorimotor loops (*reflex-like, reactive* and *anticipatory*). These conceptual results are discussed in chapter 12, which evaluates the combined results from chapters 9-11 in the light of the larger question of the sensorimotor basis of time and space outlined in chapter 8. It sketches out hypotheses and ideas for further interdisciplinary research to advance and confirm the theoretical insights gained from the current study. The last chapter 13 evaluates the methodological significance of both the simulation models presented in earlier chapters (4-7) and of the interdisciplinary approach taken in this second part of the dissertation.

Section 10.1 of this chapter describes technical aspects of the simulation model; section 10.2 presents its results. From these results, a number of interesting hypotheses have been derived, which are summarised in section 10.3. Again, the terms 'empirical' and 'experiment' are reserved for the real world experiments with humans, while the terms 'simulated', 'synthetic' and 'model' are used to refer to the ER simulation of the task and its results.

10.1 Model

The virtual task environment, in which the agents were evolved is in most respects identical to the one used for the real experiment (described in section 9.1), i.e., artificial agents can act by moving left or right in an infinite one-dimensional space (see figure 9.2 (A)), while evenly spaced objects

(same sizes, distances, velocities etc. as in the experiment) fall down in a direction vertical to the agent movement and have to be caught. Each trial consists of a sequence of 32 objects at variable random velocities (i.e., the agents were not tested on the fixed sequences across conditions that the participants were tested on). Even though the size of the agent's perceptive field is the same as the human participants' (16×8 units), I decided not to model the exact tactile input patterns the participants received, but to simply feed a continuous input signal representing the horizontal distance from the centre when an object entered the receptive field ($I_1 = |d_h|/6$ if $|d_h| \leq 6 \wedge d_v \leq 16$). The auditory pulsed signals to indicate the velocity of falling objects used in the experiments were fed into a second input neuron (I_2). The third input signal used I_3 is the reward signal, in case an object is caught (rectangular input for 100 ms). Just as in the experiment, an object is caught if it is in the centre region of the agent's receptive field when reaching the bottom line ($|d_h| < 4 \wedge d_v = 0$).

All three input signals are fed into the control network scaled by the sensory gain S_G and with a temporal delay. As explained in section 9.1.2, in the condition 'without delay', there is a minimal processing delay (on average 35ms) in the experiment, which is prolonged by 250 ms to 285 ms in the delay condition. The same values (i.e., 35 and 285ms) are used in the simulation. The agents are controlled by a CTRNN (cf. equation 3.2 in chapter 3). Three input neurons feed forward into a fully connected layer of six hidden neurons, which feed the two non-recursively coupled output neurons. I chose a time step of 7 ms for both the simulation of the network dynamics and the task dynamics, which is a much better temporal resolution for the simulated environment than in the real experiment (ca. 15 ms), in which temporal exactness was limited by the necessity for the simulation to run in real time with concurrent processes, such as reading input, displaying tactile output and writing the data to a file. The basic velocity output v calculated by the network is $v = \text{sign}(\sigma(a_{M1}) - 0.5) \cdot M_G \cdot \sigma(a_{M2})$, so one neuron controls velocity and another one direction, the motor gain M_G scales the output.

The search algorithm used to evolve the parameters of the control network is the standard generational GA described in section 3.3, vector mutation of magnitude $r = 0.6$ was used. The parameters evolved (145 parameters) are: $S_G \in [1,50]$, $M_G \in [0.001,0.1]$, $\tau_i \in [25,2000]$, $\theta_i \in [-3,3]$ and $w_{i,j} \in [-6,6]$. The fitness $F(i)$ of an individual i in each trial is given by the proportion of objects caught

$$F(i) = \frac{1}{32}\sum_{1}^{32} d_{hi}(T) < 4 \tag{10.1}$$

which is equivalent to the performance criterion used in the experiment (equation (9.1)).

I evolved agents to solve the task both with and without delays. Initially, this was intended as just the first step for a series of simulation models, with the ultimate goal to evolve agents with the capacity to adapt during their lifetime to lengthening or shortening of the delays. Since, however the agents evolved in the majority produced no negative after-effect for shortening of delays, there was no selection pressure to evolve more interesting mechanisms of adaptivity than just this robustness. Also, the robustness of solutions to the delay scenario was something the artificial agents shared with the human experimental participants. Since this simulation experiment was not primarily intended as a theoretical study of the principles of adaptation to sensory delays but as a model of the empirical experiment, limited adaptivity and sophistication of evolved solutions was actually a good thing, because it mirrored the problems I had faced in the real experiment.

10.2 Results

From 10 evolutionary runs with 1000 generations for each condition, I had to discard one from each in which simply nothing evolved. Otherwise, solutions for both conditions evolved to a high level of performance (see figure 10.1).

On the level of behavioural strategy, the solutions evolved for both scenarios involve halting abruptly once the object is encountered, frequently slightly overshooting the target, to then invert velocity and slowly move back to place the object in the centre of the receptive field. Figure 10.3 (A) shows how this strategy, from different starting positions relative to the object, leads to a stabilisation of position by performing a temporally displaced stereotyped movement. This is a rather trivial strategy. It is probably due to tight temporal constraints on the task and the coarseness of the fitness function, that does not well capture the subtleties of sensorimotor perturbation and adaptation and thus does not encourage the evolution of adaptive or more variable behaviour (see following analysis).

Figure 10.1 displays performance across the four conditions (evolved/tested with/without delay), which I see as metaphors for the conditions pre-test, delay-test, adaptation-test and post-test from the experiment. Comparing figure 10.1 with figure 9.4 from the experiment, it is striking that the evolved results are much less variable than the experimental results. Another difference is that evolution with delays achieves a much higher performance (similar level as without delays) than participants' performance in the adaptation-test of the experiment.

Comparing how agent performance and changes with introduction/removal of the delay, it is obvious that most of the solutions to the task with sensory delays are robust to the removal of the delay, while most of the solutions evolved without delays suffer a drastic breakdown in performance below chance level once the delay is introduced. This result is, to a degree, analogous to the experimental data, in which the delay condition was characterised by a catastrophic performance break-down, whereas removal of the delay led to the immediate recovery of original performance levels. If solving the task with delays in many cases subsumes solving it without in the given experimental set-up, we would have a very simple explanation for the fact that no negative after-effect could be measured in the experiment.

A closer look at the solutions evolved reveals that the velocity at which the object is first touched is on average twice as high for the controllers evolved without delays ($\bar{v} = 0.025$) than it is for the controllers evolved with delays ($\bar{v} = 0.014$). This difference suggested that the agents evolved may simply use the same strategy for both solutions, but slowing down their movement for the delay condition, which is exactly the strategy I had wanted to make impossible by including the strict time pressure in the simulation/experimental task. As argued in section 8.3, slowing down to compensate for a delay appears to interfere with semi-permanent adaptation. A very crude test of this is to invert the M_G in agents evolved for either condition, i.e., to double it for agents evolved without and divide it by two for agents evolved without delays and investigate the effect of this inversion on performance on either condition. Figure 10.2 (B) depicts the performance profile of agents upon this modification of velocities, and they seem to confirm the apprehension: with this modification, the agents evolved without delays become generalists that perform alright under both conditions, whereas the agents evolved with delays, if sped up, lose their capacity to deal with delays but are still able to solve it without delays. Halving or doubling the velocity inverts

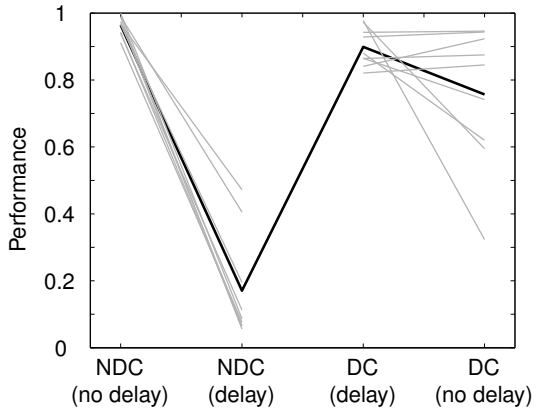

Figure 10.1: Evolved agents' performance in the condition they were evolved in and upon introduction/removal of the delay. *NDC*: evolved without delays; *DC*: evolved with delays. Black line: Average performance; grey lines: Individual fitness levels between conditions.

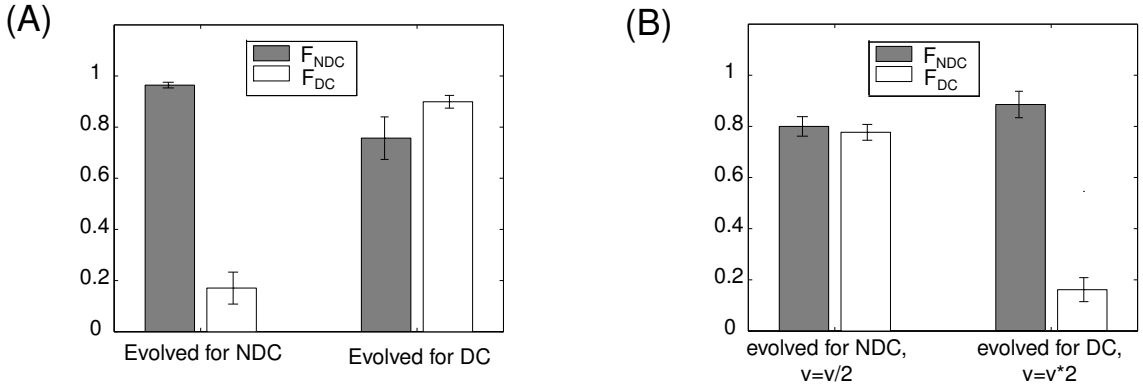

Figure 10.2: Performance profile averaged over 9 evolutionary runs in an unperturbed condition as opposed to perturbation through scaling the velocity. (A) Unperturbed condition (as in figure 10.1). (B) Scaled velocities (doubled for *DC*, divided by two for *NDC*) leads to an inversion of the performance profiles.

the performance profile evolved for each agent originally (figure 10.2 (A) or figure 10.1).

A closer look into the sensorimotor dynamics, however, shows that things are not quite this simple as the following detailed analysis of evolved behaviour shows. As a first step into the analysis, it is established that all evolved controllers function independently of the reward signal and the pace at which the objects fall (I_2 and I_3), agents simply try to put objects as quickly as possible into the centre of the perceptive field. Therefore, agents produce the same trajectories for different object velocities that are just cut off at different points in time. This simplifies analysis immensely, because object velocities can be largely ignored in the analysis and predicts a similar independence in the experimental data.

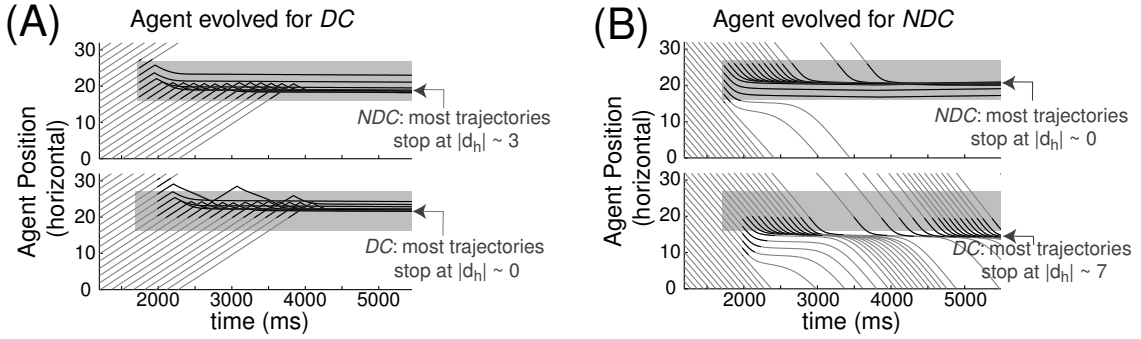

Figure 10.3: Trajectories for different agent starting positions across time, presentation of a single object. Crossing the object (grey region) produces a (delayed) input stimulus I_1 (trajectories black during stimulation). Top: without delay, bottom: with delay. (A) An agent evolved with delays. (B) An agent evolved without delays.

10.2.1 Systematic Displacements

Probably the most important result from the analysis is the identification of systematic displacements depending on initial movement direction and velocity. Figure 10.3 depicts trajectories from different starting positions relative to the object position for two example individual agents, one evolved with delays (A) and one evolved without delays (B). The agents were tested without delay (top) and with delay (bottom).

Both achieve to locate the object in the centre of their receptive field for most possible starting positions in the respective condition they have been evolved for (bottom left for agent evolved with delays, top right for agent evolved without delays). Comparing, in contrast, how the behaviour is altered by the introduction/removal of a delay (top left for agent evolved with delays, bottom right for agent evolved without delays), it can be seen that, in both cases, the trajectories are systematically displaced from the centre of the perceptive field. When the agent evolved without delays is exposed to a prolonged delay (bottom right) it overshoots its goal, while the agent evolved with delays stops too early if the delay is removed (top left). These systematicities is much closer to the behaviour I had predicted to occur in the experimental participants because both agents appear to be perturbed in their performance by alteration of sensorimotor latencies and one perturbation is the qualitative inversion of the other (negative after-effect).

Why is this systematic displacement disastrous to fitness in agents evolved without delays but interferes little with fitness of agents evolved with delays? As remarked earlier, agents evolved without delays move on average twice as fast. The magnitude of systematic displacements of the type described is proportional to the agents' velocities. The systematic displacement in the slow agent evolved with delay is small enough ($|d_h| < 4$) to still be close enough to the centre to be registered as success, as defined in the fitness function equation (10.1). For the agent evolved without delays, the displacement takes trajectories far away from the centre and outside its receptive field, as a direct consequence of the movement velocity when the object is sensed. Such systematic displacement of trajectories can be observed for most agents. The fitness function does not detect or punish such micro displacements. This appears to explain their robustness towards removal of the delay, which thus appears not to stem from a qualitative difference between removal of delay

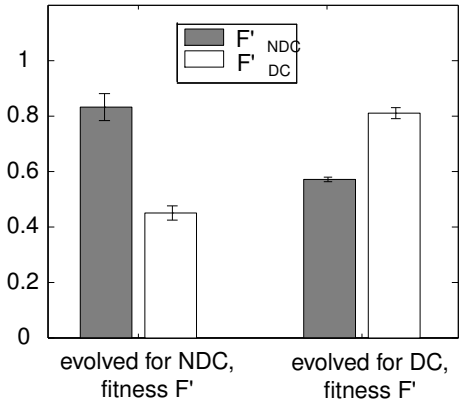

Figure 10.4: Performance profile with the modified fitness function F' (50% performance chance level).

as opposed to introduction of delay.

In order to test this hypothesis, I evolved a new set of agents with a spatially more exact fitness function

$$F'(i) = \frac{1}{32}\sum_{1}^{32} 1 - \frac{\sqrt{d_{hi}(T)}}{4} \tag{10.2}$$

With this modification, solutions evolved with sensory delays cease to be robust to the removal of the delay (see figure 10.4), which confirms that robustness of agents evolved with delays is related to the fact that the original fitness function (10.1) is immune to micro displacements. Applying this synthetic insight to the experimental study, which has the same coarse performance criterion, the model generates a possible explanation of why no negative after-effect not produced: a variable to be investigated is the displacement from the exact centre of the agent to look for systematic displacements of the described type.

This modified fitness function also allows to investigate lifetime adaptation to delays by evolving agents with variable length delay (in the original condition, the agents evolved with delays were already robust to the removal of the delay, such that there was no selection pressure to evolve more elaborate mechanisms of adaptation). However, despite longer evolution, no adaptive behaviour to adjust strategies evolved, only fixed strategies that compromise between the two conditions. As this part of the evolution is just a theoretical exercise without relevance to the experimental study, I decided not to further pursue this issue.

10.2.2 Stereotyped Trajectories

Another interesting observation about the solutions evolved is the predominance of *reflex–like behaviour*. Looking at the steady state velocities for varying I_1 representing distance from the exact centre in evolved agents (figure 10.5), there is a strong tendency to output $v^* = 0$ for values of I_1 that exceed a certain rather low threshold value of I_1. Behaviourally, this means that the agents are only sensitive to the onset of the stimulation when an object enters the receptive field, which triggers a rapid decay of v to 0. The exact magnitude of the input signal that represents the exact distance from the centre is not used for further adjustments. The variation in signal

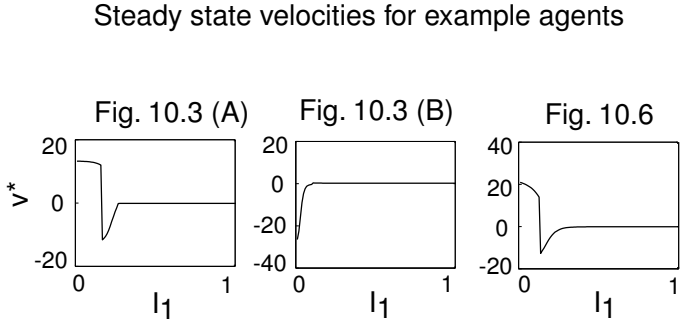

Figure 10.5: Steady state velocities v^* for different I_1 for the analysed evolved agents in figure 10.3 (A) and (B) and figure 10.6.

magnitude, as an agent moves to the exact position to stop, however, is without effect on agent behaviour. This is why the agent depicted in figure 10.3 (A), top, remains in its location displaced from the centre of the receptive field, rather than to actively search for the exact centre. Such strategies are reflex-like in that they produce stereotyped trajectories.

A variation of this pattern frequently found is that deceleration is preceded by a movement direction inversion realised by negative peaks in the steady state profile: the negative peaks in v^* in figure 10.5 (left and right) realise this return behaviour (cf. figure 10.3 (A)). Such return strategies are, however, equally insensitive for exact signal magnitude.

Reflex-like behaviour evolved in all agents but one. The agent evolved without delays whose behaviour is depicted in figure 10.6 is one of the two agents that maintain a relatively high level of performance when sensory delays are introduced (cf. figure 10.1). Even though the strategy evolved is also reflex-like in its 'native condition' (i.e., without delay), it allows adjustment of behaviour to a certain degree after performing the first reflex-like positioning: crossing the object, the target is overshot by a large amount and the first movement inversion (induced by lower negative peak in the steady state velocity profile in figure 10.5 right) positions the object in the centre in the condition without delay. In the condition with delay, however, this reflex happens to bring the object back into the outside margin of the perceptive field where the other negative peak in the steady state velocity profile is situated (figure 10.5 right). Therefore, another return reflex is triggered that brings the trajectory into the centre. In this sense, the behaviour is more *reactive*, because it is sensitive to changes in magnitude of the signal caused by ongoing behavioural dynamics (figure 10.6 top vs. bottom).

This reactive strategy is, however, plainly accidental and not the outcome of artificial evolution: if the magnitude of the return trajectory or the initial velocity were a bit different, the second inversion of velocity would not be realised in the delay condition that the agent was not evolved on. Reactive strategies did not evolve systematically because the deliberate inherent time pressure in the task does not allow for them to be beneficial. The cut off time for trials with the top three velocities is 1000, 1142 and 1333ms after the objects become perceptible, which corresponds to the vertical lines at $t = 2701, 2843$ and 3033 in figure 10.6. A reactive mechanism to bring back overshooting trajectories needs more time to come into effect. The time window is just big enough to execute a reflex, not for reactive correction, and agents have to induce the right behaviour im-

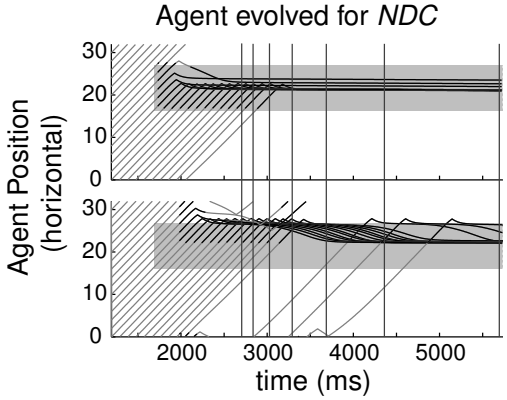

Figure 10.6: Trajectories for different agent starting positions across time, presentation of a single object. The agent that has been evolved without delays uses a reactive sensorimotor strategy. Crossing the object (grey region) produces a (delayed) input stimulus I_1 (trajectories black during stimulation). Top: without delay; bottom: with delay. Vertical lines: time at which presentation is cut off depending on v_o.

mediately when the object is perceived.

10.2.3 Velocity

The question remaining is why the solutions evolved for the task without delays are so much faster than those evolved without. The intuitive answer to this question is the wrong answer: slowing down seems the obvious way of coping with a delay - this intuition is, however, only directly true for reactive strategies, in which ongoing behaviour correction is informed by and has to wait for the delayed signal representing the effect of one's own previous actions. For the execution of a reflex there is no immediate disadvantage to high velocity giving a situation with delayed sensation. I investigated three other possible explanations.

Firstly, I hypothesised that may be correlated to maximum periods of drifting back, realised through the combined use of fast time constants in the direction neuron, and slow time constants in the velocity neuron - however, looking at the τs evolved shows that there is a general trend towards minimal τ for both motor neurons in controllers evolved for both conditions.

The second hypothesis was that the minimal reaction time t_r in the task is a function of the sensory delay $t_r(d) = t_n + d$ (where t_n is the controller-internal reaction time) and that the networks would optimise velocity in order to use this minimal reaction time to localise the centre of the object (6 units). Were this the case, v should be such that $t_n = 6/v - d$ is near constant across evolved networks. Calculating this value as a function of the evolved velocities, however, several orders of magnitude of variation between and within networks evolved for both conditions result. This means that there is a lot of variation as regards the time occupied to arrive at the centre, and that selection pressure does not operate to optimise velocities in the described way.

The third and last hypothesis tested was that the shortening of the absolute time window in which to solve the task by 250 ms in the trials with delay makes a difference and gives the networks evolved without delay more freedom to deviate further from the centre before focusing. However, testing the networks evolved without delay with faster object velocities to compensate for this

difference in time window led only to a marginal (5.6%) decrease in performance. I am still unsure why faster solutions evolved in the condition without delays. There seems to be no simple answer, even if the answer may well be a combination of several of these simple factors investigated.

10.3 Discussion

The model generates a number of insights into the task and the constraints it imposes on the strategy space, which lead to predictions about descriptive properties of the data obtained from the experiment presented in chapter 9. The following chapter re-analyses the experimental data to test some of these predictions resulting from the simulation model.

Most noticeably, the model shows how the coarse fitness function does not register the subtle systematic displacements that result from the removal of sensory delays as unsuccessful behaviour. These displacements can be seen as the analogy of a measurable after-effect in the experiment. Movement velocity could be shown to play a role in explaining differences in displacement magnitude. In this sense, the model predicts that the participants in the experiment *overshoot* their target when the delay is introduced, that this overshooting decreases over training, and that they *stop earlier* when the delay is removed. Finding this kind of regularity could give evidence for the correctness of the original ideas underlying the main hypothesis, even if the hypothesis is not confirmed by the experimental results. As part of the findings on systematic displacements, the model predicts that velocities decrease between pre- and post-test.

A third issue the simulation suggests for analysis is the reflex-likeness of behaviour. From the simulation we expect that, since a delay prolongs the absolute temporal duration of a closure of the sensorimotor loop, the strategies become more reflex-like over training with delays. There are many possibilities of how reflex-likeness could be defined and measured. A simple measure explored in the following chapter is the intra-participant-similarity of trajectories. This prediction is relevant in arguing the theoretical results from the interdisciplinary study that are developed in chapter 12. The simulation also predicts that systematic displacements should be more pronounced in strategies identified as reflex-like.

The evolved agent controllers have very simple strategies that rely on only few sensorimotor invariants. Factors that do not matter to evolved strategies are the velocity of the objects (catch as fast as you can), the history of previous object presentations, the exact magnitude of the tactile input and the auditory reward signal. The model predicts that the participants' strategies are similarly independent of these factors.

Many of these predictions are tested in the following chapter, and some of them are supported by the data. In principle, the possibilities for further analysis of experimental and simulation data are open-ended. Within an interdisciplinary paradigm, however, there is a point where further experimentation is more promising than further theorising, and I believe the analysis performed here may have already been a bit too detailed given the original hypothesis, the fact that it was not supported and the problems identified so far. However, as this is the first application of the proposed interdisciplinary method, it appears reasonable to fathom out all possibilities, in order to gain an impression of the methodological potential of the combination of ER and PS research developed. Chapter 13 performs such an analysis and identifies space for improvements, despite an altogether positive conclusion from the work presented.

Chapter 11

Further Data Analysis in the Light of the Simulation Results

In this chapter, the experimental results from the experiment described in chapter 9 are revisited in the light of the simulation results. The descriptive concepts and variables pointed out by the simulation model presented in chapter 10 are investigated in order to yield a more detailed description of the data and to quantify in which way it had been transformed over training with delays.

The data had to be re-ordered and filtered (section 11.1). Section 11.2 investigates the different dimensions that were predicted in the simulation model to be relevant. In particular, subsection 11.2.1 investigates systematic displacements in the experimental data. Subsection 11.2.2 takes a look at the initial movement velocity of participants and subsection 11.2.3 explores intra-participant similarity of trajectories as an approximation of how stereotyped and reflex-like the strategies are. In subsection 11.2.4, the issue of inter-participant variability in strategies is discussed, before, finally, in subsection 11.2.5 it is investigated whether object velocities play a role in the experiment. The results obtained are briefly summarised in section 11.3.

Chapter 12 discusses and evaluates the results and conceptual analyses presented in this chapter and the previous chapters 8-10. The conclusion chapter 13, finally, evaluates the methodological success of applying the interdisciplinary framework developed in chapter 3 to the problem of adaptation to sensory delays and concludes that, through the different modelling approaches taken in this dissertation, the explanatory potential of ER modelling for human Cognitive Science has been confirmed.

11.1 Pre-processing

In order to be able to investigate the issue of systematic displacements and stereotyped trajectories, the data has to be restructured. In its raw form, it is simply a long sequence of positions, corresponding time tags and states of sensory stimulation. As the variables predicted to be relevant are independent of the order of object presentations, the sensorimotor data of each participant for each of the four experimental phases compared (pre-test, delay-test, adaptation-test, post-test)

was segmented into 32 object presentations associated with each condition.[1] This 'bag of 32 catch intents' per participant and condition contained catch intents for objects of different velocities. As the results obtained from the simulation model were independent of object velocity, testing these predictions allows this generalisation. As the following analysis is mainly based on this pre-processing, the role of object velocity is not taken into consideration. However, section 11.2.5 explores whether or not some essential variables depend on object velocity. Out of the 2560 catch intents classified (20 participants $\times$ 4 conditions $\times$ 32 objects), 198 (ca. 8%) were immediately removed, because they either did not involve any established contact with the presented object or because the participant remained completely immobile.

The second stage of the filtering and restructuring consisted in separating intents of catching objects from the right from intents to catch objects from the left. For analysing sensorimotor data from human participants, differences in direction crucially matter, due to the physiological asymmetry in moving a mouse left or right - some participants' behavioural strategies were qualitatively (e.g., strategy) or quantitatively (e.g., initial velocity) contingent on movement direction (see, e.g., behaviour by participant (A) depicted in figure 11.1, left).

Left and right catching intents were separated and treat both sets as if they were from two different participants, one, who exclusively searches for objects to the left and one who exclusively searches to the right. The formal criterion for this separation was the movement direction of participants before they made contact with an object. This segmentation leads to 160 classes of catch intents (20 participants $\times$ 4 conditions $\times$ 2 directions).

This organisation has the additional advantage, that the direction in which to expect systematic displacements, is normalised, at least in so far as the participants apply the same strategy across catch intents from one direction, an issue that becomes relevant again in section 11.2.1.

Unlike the simulated agents, participants use both directions alternatively, even if there was an average directional preference of 63% (max. 86%). Directional preference in participants also changes between the pre- and the post-test by an average of 19%. By treating left- and right catching attempts separately, such interesting aspects of the data are lost.

The next and most complicated step of the data preprocessing consisted in calculating mean trajectories for each of these collections of intents. This calculation of mean trajectories has two advantages: firstly, it provides formal grounds for eliminating outlier catch intents for participants who otherwise behave very regularly on the task: in some cases, behaviour was irregular (e.g., slips of the mouse, distractions, performance issues with the system) and induces noise in the data. Investigating intra-participant-similarity of trajectories, such outliers can be eliminated (see below). Also, the same measure can be applied as a measure for stereotypedness of trajectories, testing one of the descriptive predictions generated from the simulation model (section 11.2.3).

In order to compute average trajectories, the recordings from each object presentation per subject and movement direction were normalised in time in space, defining the first point of contact with the object (not the point where objects were in principle visible!) as the point of reference. In the resulting normalised data, this corresponds to 1640 ms time and 0 space (see figure 11.1). With this choice, the trajectories for a perfectly stereotyped reflex-like behaviour would have been exactly congruent after this transformation.

[1] The reader is reminded that the performance is specified as success on a fixed sequence of 32 objects presented (cf. chapter 9, section 9.1.1 and equation (9.1)).

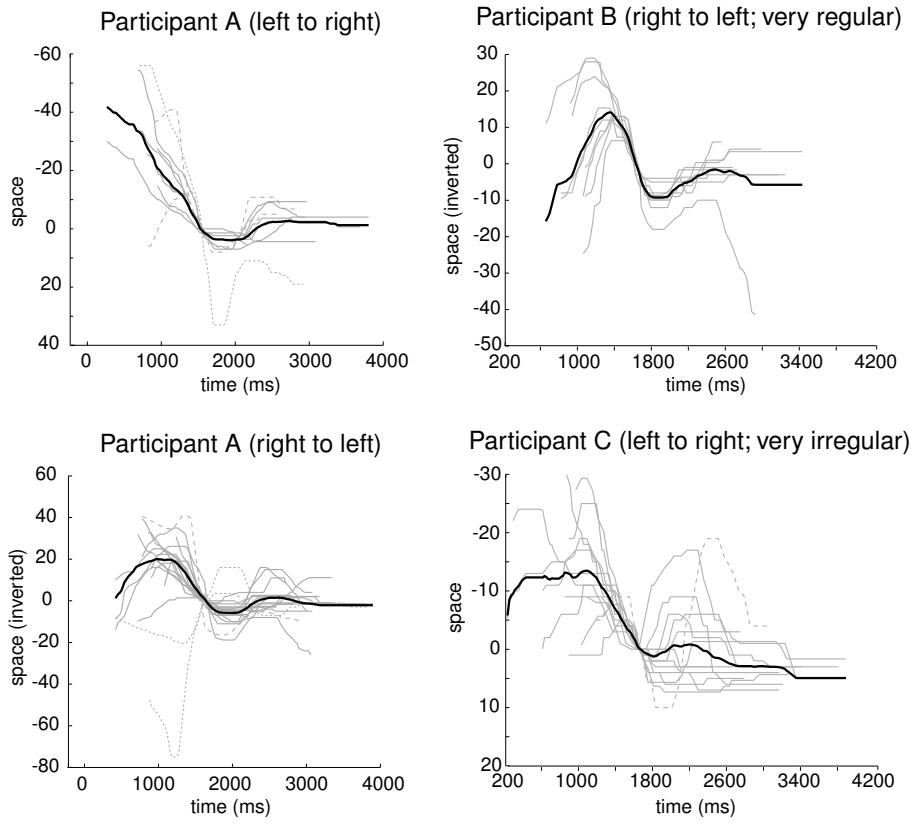

Figure 11.1: Examples of sorted and superimposed intents to catch objects during the pre-test phase. Black: average trajectory; grey: individual catch attempts; dotted: removed by filter ($> 3\sigma$ from mean trajectory). Left: a participant (A) whose left scanning movements (bottom) are qualitatively different from her right scanning movements (top). Right: a participant (B) with very stereotyped trajectories (top) and one (C) with very variable behaviour (bottom).

In order to eliminate temporal irregularities in frequency of written data points due to system limitations (cf. section 9.1.2), I normalised and regularised the time line into segments of 20 ms, filling in the gaps assuming that the positions change linearly and continuously between actual existing data points. Due to differences in stimulation onset and object velocity, trajectories did not have unified starting and end points. I discarded data from intervals in which less than four trajectories were recorded. I then calculated the average trajectories by averaging the change in position at any point.

Those intents were defined as outliers whose mean square difference in change of position at any data point was more than three standard deviations (σ) away from the average trajectory. As the original mean trajectory included these outliers, which sometimes had strong effects on the average trajectories and on σ, the average trajectories and σ were recalculated after the first removal of abnormal trajectories. The filter was re-applied after recalculating the average. After three repetitions of this cycle, the average trajectories had stabilised and nearly all of the trajectories that intuitively appeared to be outliers had been removed. Altogether, another 157 trajectories ($\approx 6\%$) were thus removed (see figure 11.1; dotted trajectories were classified as outliers and removed).

Again, the classification and averaging of trajectories had relied on the assumption that the data recorded from the mouse was actual position data. The application of the acceleration curve by the operating system (cf. chapter 9 section 9.1.1) implies that averaging the distance in position and filtering on the basis of σ is not actually formally justified. However, the results are still presented here a) because they provide a good example of how the proposed framework can work in general and b) because they still provide a valid description of what can be expected to happen in a formally correct version of the experiment, as the effect of the acceleration curve within a certain velocity band is fairly subtle.

My opinion is that the acceleration curve may have played a negative role in identifying wrongly an outlier (when it is in fact an accelerated version of the common strategy used in the other cases), but those that are contained within the distribution are unlikely to be false positives (of course not impossible to think of the acceleration curve distorting a trajectory so that it looks like an example of this strategy while in fact it is an example of that one, but I think unlikely).

The last step of the filtering was to fully remove one scanning direction for one participant, because it was too sparse to meaningfully analyse the average behaviour (less than 5 intents during pre- and post-test after filtering). This corresponded to another 29 trajectories being removed, which is just over 1%). Altogether, after this process, each collection of the 156 catch-intents (39 participants/direction $\times$ 4 phases) contained on average 14 trajectories, and none contained less than five.

In the simulation model, agents evolved apply the same strategy regardless of object velocity and previous object presentations. Sorting the data this way presumes that the same holds for the experimental participants within each experimental phase. There appear to be some cases in which this assumption seems unjustified (figure 11.1, participant (C) bottom right). In general, however, it appears to be surprisingly accurate, as parts of the following analysis confirm.

Table 11.1: Average value across 39 classes of intents for several variables that describe the trajectories.

	pre-test	adaptation-test	delay-test	post-test
1.) Systematic Displacements (1)	-0.1002	0.3125	0.0849	-0.2756
2.) Systematic Displacements (2)	-0.1851	0.5705	-0.1274	-0.1418
3.) Systematic Displacements (3)	-0.2025	0.6412	-0.0289	-0.3032
4.) Velocity during scanning	0.0330	0.0266	0.0280	0.0295
5.) Velocity last 500 ms before touch	0.0384	0.0324	0.0290	0.0321
6.) Velocity after touch	0.0032	0.0071	0.0044	0.0032
7.) Velocity after touch whilst mobile	0.0180	0.0225	0.0176	0.0197
8.) MSE from average trajectory	0.4267	0.7229	0.3549	0.2892

11.2 Statistical Analysis of Filtered Data

The filtering described in the previous section compiled the data into a format that allows to test the main hypotheses generated from the computational model. The findings presented in this section focus on the participants motion across time, not sensory input. Sensory inputs are only taken into consideration in so far as the onset of sensation is the point of reference for normalisation of data in the temporal domain.

In a full dynamical analysis, investigating how motion and sensation in experimental data relate over time is essential. The possibilities for such analyses these are open ended, starting from simple cross-correlation measures to more sophisticated relational measures such as Granger causality (e.g., Seth & Edelman, 2007). Cross-correlation of sensation and motion in the data gathered was investigated to a certain extent. However, the results, even though potentially interesting, would have required further work in order to unambiguously support a point. Putting more work into analysis did, however, not appear reasonable, given the fact that the experimental results are problematic and do not support strong claims about the sensorimotor basis of experienced simultaneity. In principle, such data-driven dynamical analysis is a good and important complement for the more model-driven analysis presented in the following sections.

The variables looked at in more detail as to how they develop across the different experimental phases are: the systematic displacements upon catching the object (subsection 11.2.1), the velocity before touching the object (subsection 11.2.2) and the mean square error from the average trajectory as a first approximation of how stereotyped trajectories are (subsection 11.2.3; see table 11.1 for a summary of the values investigated. Further on, the issue of inter-participant differences in strategy (subsection 11.2.4) and the role of object velocities (subsection 11.2.5) are explored.

As mentioned in section 11.1, unintended mouse acceleration may have influenced the filtering preceding this analysis. Similarly, the assumption of linearity of the measured data is not fully justified for the same reason.

11.2.1 Systematic Displacements

The main prediction from the simulation model is that a clear negative after-effect occurs in terms of changes in systematic relative displacements between the centre of the receptive field and the centre of the object to be caught at the end of an object presentation that depend on initial movement direction and velocity. Due to the coarseness of the performance criterion (equation (9.1)), if such systematic displacements are small enough in magnitude, they are not necessarily reflected in catch-performance, which could explain the negation of the main hypothesis.

Findings on systematic displacements presented in this subsection are ambiguous and do, altogether, not support the prediction generated from the simulation model. The first discrepancy between the model and the experimental results is the initial direction of change. In the simulation model, the time window in which an object could be centred was so short that the evolved agents performed a reflex-like trajectory to reach their final position. This implies that systematic displacements were always in direction of initial movement when the delay was introduced and always in the opposite direction when it was removed. Investigating the direction of change in displacement between the pre-test and the delay-test, it was found that this is not the same for the experimental participants. In the empirical data, only 28 of the 39 collections of intents showed an average displacement in direction of the initial movement. In 10 of the remaining collections, the initial change in displacement was in the opposite direction as the initial movement direction and one did not exhibit any average change in displacement when the delay was introduced. Therefore, there was no clear basis for testing this prediction in the experimental participants.

I tested the prediction about systematic displacement applying three different criteria and obtained mixed results (table 11.1, systematic displacements (1), (2) and (3)).

Assuming that *initial movement direction* is the most important dimension to characterise systematic displacements, the displacements can be calculated as the distance d_{tt} of the receptive field at the end of an object presentation from the exact object centre *multiplied by the sign of the initial movement direction*. In this case, the displacements of the location where the participants stop to catch the object follows the pattern predicted (see figure 11.2 (A) top and table 11.1, variable 1.)). The initial displacement of $d_{tt} = -0.1$ rises to 0.31 when the delay is introduced; This perturbation is partially recovered and drops back to 0.08 over training. In the post-test, it decrease even further to -0.28 when the delay is removed. While the difference between pre- and post-test is not statistically significant ($p = 0.1852$), the *mean difference* in displacement (figure 11.2 (B) top) from pre-test to the delay-test and from the adaptation-test to the post-test is ($p = 0.14 \cdot 10^{-3}$). Additionally, they have the same magnitude ($p = 0.79$), but different directions. This seems like a clear confirmation of the negative after-effect predicted by the simulation study. It raises, however, the question why in ten participants the initial change in displacement location followed the opposite direction predicted and how these figure in the calculation of averages.

To test the role of these irregular data profiles, the data was restructured. Assuming that the *initial direction of change* in systematic displacements is the important variable (i.e., the displacements are calculated as the distance d_{tt} of the receptive field at the end of an object presentation from the exact object centre *multiplied by the sign of the change in distance between pre-test and delay-test*), the effect disappears (figure 11.2, bottom and variable 2.) in table 11.1): under this interpretation, the adaptation to delays fully restores the original average displacement, which then

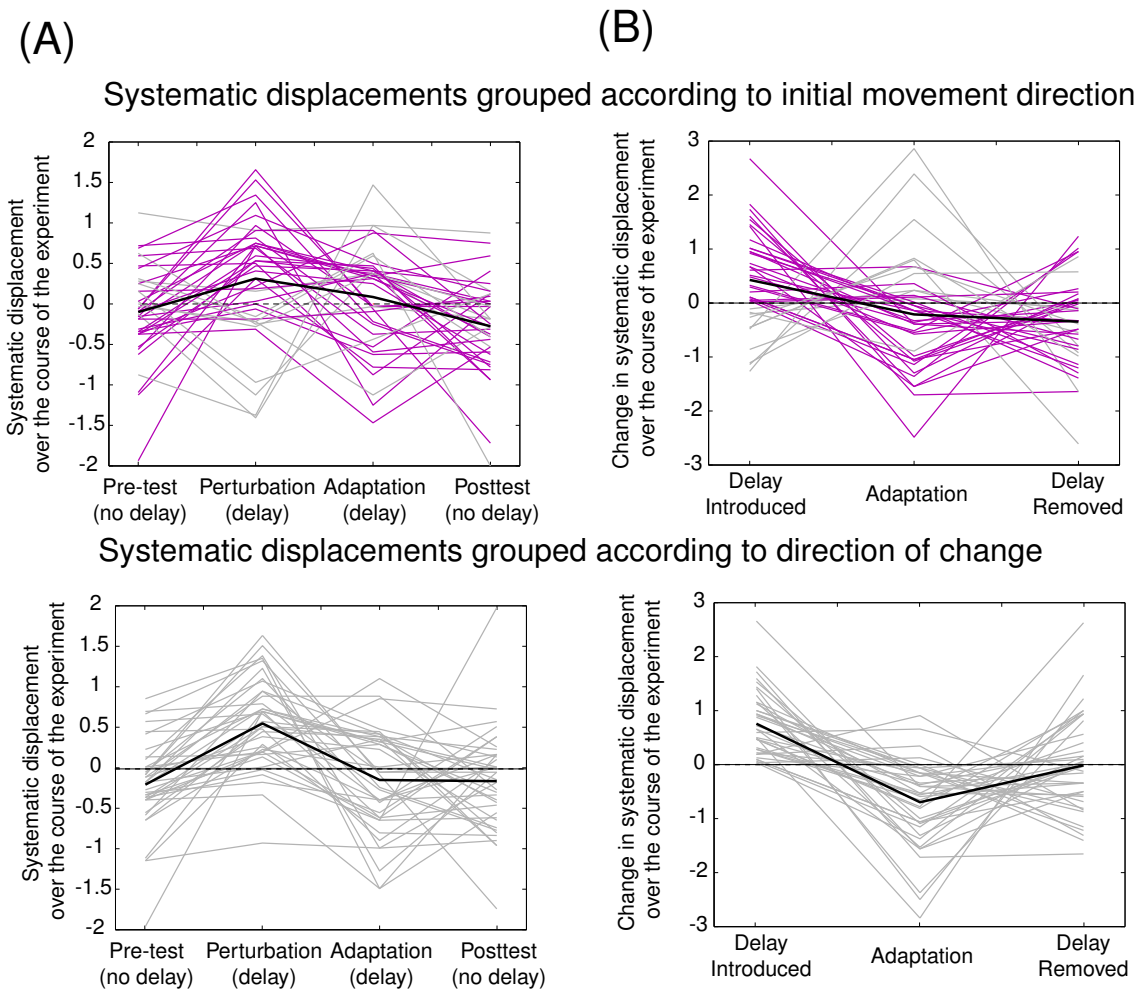

Figure 11.2: Top: systematic displacement relative to initial movement direction (systematic displacements (1)). Bottom: systematic displacements relative to the direction of change from pre- to delay-test (systematic displacements (2)). These two interpretations do not support the same conclusions. Magenta trajectories in the top graphs represent those 28 data profiles in which both interpretations agree (Systematic displacements (3)). (A) The relative displacement of the centre of the receptive field to the centre of the object to be caught. (B) the change in this relative displacement across the experimental phases.

remains unaltered when the delay is removed.

The last test of systematic displacement was to only investigate those 28 collections of catch intents, in which the *initial movement direction and the initial direction of change agree*, eliminating the eleven ambiguous data profiles. This interpretation, again, confirms the prediction that systematic displacements follow the pattern predicted by the simulation in a similar way as for initial movement direction alone (Magenta trajectories in figure 11.2, top and variable 3.) in table 11.1), even if this difference is not statistically significant.

These mixed results do not allow a clear conclusion on whether or not the prediction generated by the model about systematic displacements is confirmed, as the experimental data appears to be more variable in certain respects than the synthetic data from the evolved agents.

11.2.2 Velocity

In the evolved agents, the magnitude of systematic displacements was dependent on initial movement velocity. This difference in magnitude impacted on performance, thus explaining why no negative after-effect of removing delays was measured. The previous subsection does not support the prediction that similar regularities in magnitude of systematic displacements characterise the experimental data. In this sense, participants are different from artificial agents, and velocity differences cannot be evoked in the same sense as in the artificial agents to explain this result.

However, it is still worthwhile to take a closer look at how velocity changes over the experimental phases. As previously remarked, due to the unintended application of acceleration, the change in position which is evaluated here is not actually linearly varying velocity of the mouse, but slightly accelerated velocity interpreted in the simulation. This was, however, only detected after the analysis here presented and I, therefore, refer to the measured change in position as velocity.

The velocity analysis in the simulation had focused on initial movement velocity, not on average velocity throughout the recorded trial. In order to have a comparable measure for the experimental results, I compared the average velocity during the last 500 ms before first touching the object (see figure 11.3 (A) and table 11.1, variable 5.)). A significant ($p = 0.04$) and gradual decrease in scanning velocity between pre- and post-test was found in the data. This decrease provides evidence for a persisting adaptation effect that is not reflected in the performance profile.

Even though the raw data presented in chapter 9 already investigated movement velocities, this clear decrease in velocity had not been detected before. Velocity differences were not significant when just averaging them across the entire object presentation (cf. section 9.2). The reason for this is that the velocities during the entire phase before making contact (table 11.1, variable 4.)) and after touching the object (figure 11.3 (B) and table 11.1, variables 6.) and 7.)) is much more variable than the velocity during the 500ms before making contact. This shadows the reported statistically significant decrease in average velocity immediately before making contact. Investigating this time frame had been suggested by the simulation.

11.2.3 Stereotyped Trajectories

The inter-participant-comparison of trajectories during each phase of the experiment can be taken as one indicator for reflex-likeness of strategies, even if this measure does not make any reference

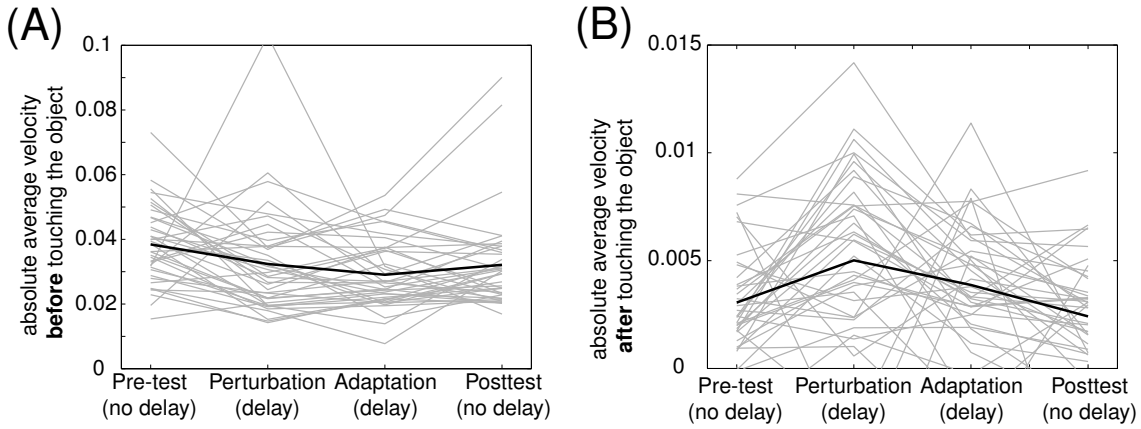

Figure 11.3: (A) The average velocity during the 500 ms before first touching the object to be caught decreases over the course of the experiment (significant distance between pre- and post-test $p = 0.04$), despite the full recovery of performance reported in chapter 9. The decrease in velocity had not been detected because of noisy velocities during other phases of the catch intent, in particular after touching the object (B).

to sensory data. Other measures that investigate the dependence on sensory signal variation are possible and interesting.[2] This subsection discusses how average intra-participant-similarity varies across the different phases of the experiment. This average is based on the mean trajectories used to filter outliers from the generated classes (section 11.1) and the mean square error (MSE) of trajectories from this value. It is important to recall that the mean trajectories and σ_{MSE} are based on changes in position which do not relate linearly to changes in mouse position.

Testing the probability distribution of this variable, it turns out that it is log-normally distributed, not normally distributed like the other variables (Jarque-Bera Test: pre-test $p = 0.3757$; delay-test: $p = 0.2756$; adaptation-test $p = 0.8454$; post-test $p = 0.9623$; Lilliefors test also supported log-normal distribution). Therefore, the logarithm of the MSE is investigated.

Interestingly, looking at the evolution of the logarithm of the MSE from the average trajectory over the course of the experiment, there is a clear decay ($p = 0.0026$) between pre- and post-test (see figure 11.4 and table 11.1, variable 8.)). This increased stereotypedness of trajectories occurs between the introduction of the delay and the adaptation to the delay and seems to be another very clear lasting effect of the training with delays.

Such effects were predicted by the simulation model, which showed that the time pressure in the task encourages reflex-like movement trajectory, not reactive adjustment of behaviour. This is already true in the case of immediate sensory delays, but even more so with sensory delays, which suggests that adaptation to delays renders strategies even more rigid.

11.2.4 Inter-Individual Differences and Strategy

The data presented in this chapter has been exclusively based on the inter-participant-average of intra-participant averages of variables. As reported in chapter 9, there appeared to be substantial variations in between participants concerning their strategy and how they react to the introduction

[2] As mentioned earlier, cross-correlation has been explored in this endeavour and produced ambiguous results.

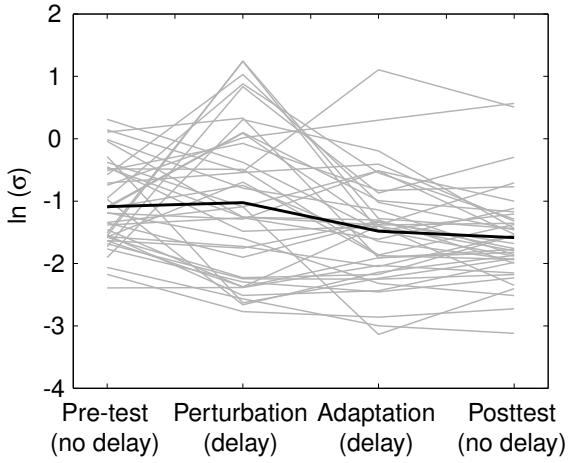

Figure 11.4: The standard deviation σ of trajectories from the mean trajectory is log-normally distributed and its logarithm declines significantly ($p = 0.0026$) from pre- to post-test. This change towards more stereotyped behaviour is initiated during training with delays.

or removal of a sensory delay. These inter-individual differences have not been taken into consideration in the previous analysis which averages variables across participants, with the exception that data had been separated into left and right catching intents. This section describes some of the attempts to fill in this gap.

The problem with inter-individual differences is the sparseness of data. Whereas in the across individual comparisons, for each phase, there were 39 data points (20 participants $\times$ 2 directions - 1 eliminated; see section 11.1), the average number of catch intents in these collection was 14, the minimum five. Therefore, applying statistical distribution tests on the variables investigated within participants and between experimental phases produced rather random and meaningless results, particularly as concerns velocity and systematic displacements. Concerning the change in the logarithm of the MSE from the average trajectory within a participant, it can be found that 46% exhibit a statistically significant decrease in the logarithm of MSE between the pre- and the post-test. This further confirms the trend towards more stereotyped trajectories as a lasting effect of adaptation to sensory delays, as predicted by the simulation model.

There are several lessons to be learned from this. Firstly, in designing future experiments, a worthwhile consideration is to test the hypothesis on a variable that is easier and more reliably to quantify and provides more statistical leverage. However, even within the experimental paradigm used, a technique that has not been applied a lot is to measure variables in each phase in relation to a participants across experiment average, thereby pointing out variations relative to a participants strategy (such as variable 7.) in chapter 9, table 9.1).

11.2.5 Object Velocity

In the analysis, I assumed that certain factors that are irrelevant to the simulated evolved agents (exact signal structure, object velocity, history of movement, reward signal) are irrelevant to the participants' strategy as well. It may not be considered crucial to demonstrate that the simplifying assumptions are valid. It is, however, interesting to investigate these assumptions, as an addi-

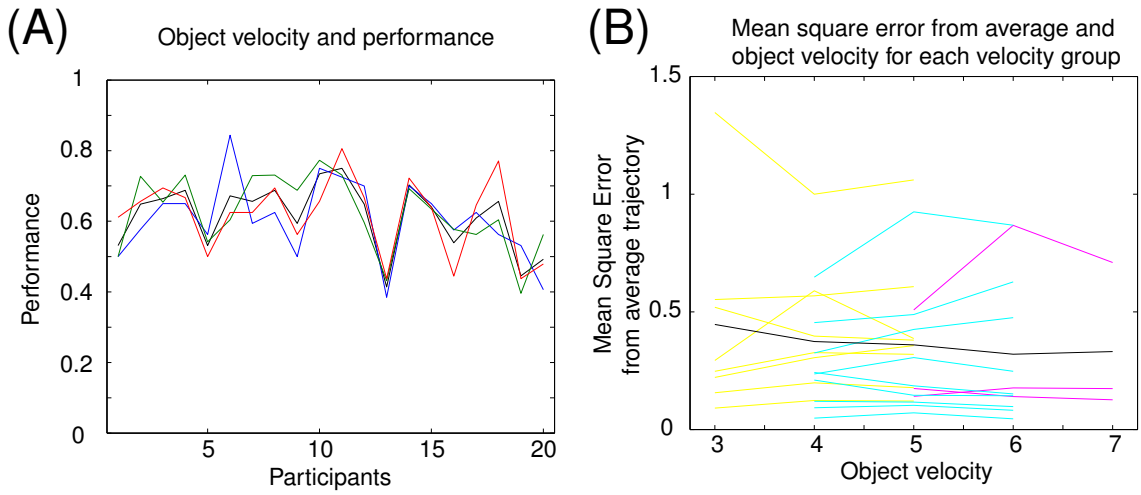

Figure 11.5: (A) Each participant's performance depending on their relative top (green), middle (red) and bottom (blue) velocity. Object velocity does not seem to be systematically linked to performance. (B) The mean square error from the mean trajectory as it varies with object velocity across the different velocity groups (group 1 (fast): magenta; group 2 (medium): cyan; group 3 (slow): yellow). Object velocity does not seem to play a crucial role for this variable

tional tests of predictions generated by the simulation model. This section investigates whether performance (as defined in equation (9.1)) or behaviour (in terms of MSE from average trajectory) is dependent on object velocity or not. The simulation model predicts that there is no such dependency.

Object velocity may be assigned particular importance, as it is variable between sub-groups of participants (classification in three velocity groups, see chapter 9, section 9.1.3), and it cannot be assumed *a priori* that this classification is irrelevant for the change in variables measured. Furthermore, given that participants within each group were presented with objects of three different velocities, it is, in principle, possible that performance or rigidity of movement vary across velocities.

Effectively, object velocity appears not to impact on behavioural success or strategy, as the model predicted. In each velocity group, participants were confronted with objects of three different velocities during the experimental phases on the basis of which performance is compared. Averaging performance across subjects and groups for their respective slow, medium and fast object velocity tested, the average values are 0.6056, 0.6233 and 0.6170; the negligible differences in performance are far from statistically relevant. Figure 11.5 (A) illustrates how, for each individual, the velocities (blue, green, red) seemingly arbitrarily correspond to the performance.

Similarly, the stereotypedness of trajectories seems to be independent of object velocity: figure 11.5 (B) shows the variation in this variable across the compared velocities 3-7. Even though, on average, there seems to be a decline in MSE from the average trajectory, looking at the different velocity groups (group 1 (fast): magenta; group 2 (medium): cyan; group 3 (slow): yellow), this seems to be rather due to different individuals dominating different velocity bands, not to a direct relation between stereotypedness and object velocity. The prediction from the simulation model

that object velocity does not impact on behavioural strategy or success seems at least not refuted.

11.3 Summary

The simulation model has generated a number of insights about the sensorimotor dynamics of the task posed to the experimental participants and the strategies afforded by the environment. The synthetic results generate a number of descriptive concepts and predictions about variables along which the sensorimotor data can be analysed in order to find empirical support some of the insights gained. This chapter presents results from data analysis of several of those variables, i.e., systematic displacements, scanning velocity, intra-subjective stereotypedness of trajectories and dependence on object velocities.

The analysis is overshadowed by the fact that the position data measured is distorted through the unintended inclusion of the operating system's acceleration curve, which renders both the data pre-processing (section 11.1) introduce bias in the pre-processing. Therefore, the results obtained can be seen more as a heuristic of the results potentially obtainable.

The results from testing the predictions generated by the model are the following.

- There was no clear evidence for the existence of systematic displacements in the sense they occurred in the simulation study. Due to inter-participant variation in strategy, measuring this variable was not as straight forward as in the case of the simulation model and different equally reasonable criteria produced opposing results.
- Concerning the movement velocity, the prediction that velocity before making contact with an object would decrease between pre- and post-test was confirmed. However, it is not clear what this decrease in velocity implies, since its functional role had not even been clear in the simulation model.
- As an approximation of the stereotypedness or reflex-likeness of strategies, the intra-subjective similarity of trajectories, measured as the $\sigma(MSE)$ from the average trajectory across catch intents, could be shown to follow the pattern predicted, i.e., to decrease significantly between pre-test and post-test.
- A final prediction generated from the simulation model and that is supported by the data analysis is that performance and sensorimotor behaviour are independent of the variation in object velocity.

Altogether, the model had been very useful both on the quantitative descriptive level (e.g., predicting decrease in velocity and $\sigma(MSE)$) and in the conceptual-generative level (e.g., role of systematic displacement with respect to velocity and motor strategy; classification of qualitatively different strategies, cf. chapter 12).

Throughout this chapter, I point out avenues for further or better data analysis, failures on my part to attend to important factors and variables and design decisions in experiment model and tools for analysis that would have yielded stronger or less ambiguous results. There were a number of things that were sub-optimal about the study on the adaptation to sensory delays. Firstly, the fact that it did not support the main hypothesis tested was, naturally, not welcomed and impacted on the enthusiasm with which data analysis was conducted. Secondly, a number of factors in the conduction of the experiment mean that they are scientifically questionable, such as the noisiness of the environment in which participants were tested, the irregularities in the

spatial properties of the environment and, most importantly, the unintended acceleration of the mouse movement data which means that the data is not linear and, thus, that none of the variables found to exhibit statistically significant variation rely on formally correct measurement. A third source of problems was the explorative nature of the project and personal lack of experience in designing and conducting experiments, which led to a number of avoidable procedural errors, such as the lack of statistical leverage concerning the performance criterion defined or the failure to record participants age and gender. Taking all these sources of interference together, the project conducted leaves me with a good idea of the things I could have done better, rather than with a good idea of adaptation to sensory delay and experienced simultaneity.

In the light of these adverse circumstances, the results obtained can be interpreted, in the first place, as a methodological contribution, i.e., a case study to test the proposed integration of ER simulation modelling into existing experimental approaches to the study of perceptual experience and sensorimotor behaviour. The scientific contribution to explain the phenomenon investigated, i.e., perceptual and sensorimotor adaptation to sensory delays, is, in comparison, secondary and of inferior quality, at least concerning the empirical experiments. This matter of fact, i.e., that mistakes were made that appear avoidable in retrospect, is probably not surprising, given the novelty of the approach taken.

The following chapter evaluates the results here presented in the context of the results presented in the scope of this interdisciplinary project, before the merits of the combined modelling and experimental approach are discussed in chapter 13.

Chapter 12

Discussing the Results on Adaptation to Sensory Delays

If chapter 8 was like a super-sized introduction section, this chapter is like a super-sized discussion section. The findings obtained from the interdisciplinary project on delays (chapters 8-11) are discussed and evaluated with respect to the hypothesis tested and also in the broader context of the sensorimotor basis of time cognition introduced in chapter 8. The following and final chapter then evaluates this project along with the previously presented simulation models (chapters 4-7) in the light of the unifying methodological research question of how ER simulation models can enrich the enactive study of human cognition and behaviour.

Section 12.1 summarises the results obtained through the experiments and the simulation (chapters 9-11) and section 12.2 evaluates them with respect to the initial hypothesis, as well as drawing the links between the findings obtained and the ideas outlined in the earlier sections of chapter 8 on time experience across disciplines and approaches. Finally, section 12.3 indicates avenues, hypotheses and plans for future research on adaptation to sensory delays and the experience of simultaneity.

12.1 Summary of the Results

I designed the described experimental paradigm that was inspired by Cunningham et al.'s (2001a) findings on semi-permanent adaptation to visual delays leading to a negative after-effect and an experiential re-adjustment of perceived simultaneity. I wanted to test the author's hypothesis that it is the inherent time pressure in the task that is responsible for this interesting adaptation effect that had failed to occur in similar earlier studies. The experimental paradigm followed the minimalist approach described in chapter 3 that is exercised by the GSP who hosted me during the experimental phase of this project. The visuomotor avoidance task used by Cunningham et al. was simplified, turned into a catch task and transferred to the audiotactile platform Tactos (Gapenne et al., 2003) in order to be able to fully control and record sensorimotor data and dynamics. The objective was to first reproduce the adaptation effect reported by Cunningham et al. in a minimised set-up and to then identify minimally different control conditions in which the effect does not occur. Thus, the sensorimotor basis of experienced simultaneity and presentness should be elucidated by describing and analysing not only the qualitative adaptation of performance, but

also the changes in sensorimotor dynamics and strategy by which it is realised.

The main hypothesis tested in the experiment was that the participants' performance profile would follow the same pattern as reported by Cunningham et al. (2001a). This is, the decrease of initial performance level upon introduction of delay, full or partial recovery over training with delays, and a decrease of performance as compared to the initial performance levels once the delay is removed. This hypothesis was not confirmed: even though the 250 ms tactile delay strongly perturbed the participants' performance and training with the delay led to a partial recovery of performance, there was no negative after-effect to the task. In so far, the experimental results are closer to earlier results in which subjects slowed down their movement to compensate logically-cognitively to imposed sensory delays, but did not exhibit a perceptual readjustment or negative after-effect (e.g., Smith & Smith, 1962). A negative after-effect would have been necessary to indicate semi-permanent adaptation to sensory delays.

Even though there was no difference in performance between the initial and final recordings without delay (pre- and post-test), a look at the sensorimotor recordings from participants' behaviour revealed that there were substantial qualitative differences in strategy and behaviour resulting from training with delays between these two otherwise identical conditions. It also appears that the experimental set-up afforded a number of possible sensorimotor strategies in order to solve the task and that different strategies were impacted differently by the introduction/training/removal of delays. It was, however, not obviously clear how to measure, quantify or describe these changes and differences or how to test whether there were systematic effects occurring across participants, and, if not, whether there were systematicities in how certain classes of sensorimotor strategy impact on these variables.

The ER simulation model of the experiment presented in chapter 10 provided some conceptual clarity about the investigated task, the behavioural strategies it affords and possible descriptive variables to describe the transformation of sensorimotor behaviour observed in the experimental participants. Evolving simple CTRNN controlled agents to perform the task with and without delays and testing them under both conditions, I found that the simple agents were robust to shortening of sensorimotor latencies, but not to lengthening. This can be seen as a pattern analogous to that observed to occur in experimental participants, i.e., a perturbation through the delay, recovery over training and failure to produce a negative after-effect. Analysis of the sensorimotor behaviour of the agents showed that, in the evolved agents, this effect did not indicate a qualitative difference between shortening or lengthening of the delays but instead a quantitative difference in perturbation effects. Robustness to shortening is a consequence of the coarseness of the performance criterion, not of immediate behavioural readjustment to the original condition. The fact that the ER simulation and the experiment use a nearly identical simulated environment makes it possible to draw strong analogies between the two and suggests the investigation of the same variables in the experimental data recorded. In particular, the simulation predicts that (1) there are systematic displacements from the average catch location both upon introduction and removal of delay; (2) these systematic displacements are in opposite directions (negative after-effect) (3) they differ in magnitude and (4) these changes in magnitude correlate with a difference in movement velocity before first touching the object. Furthermore, the simulation model suggests that trajectories would be more stereotyped after training with delay.

The variables pointed out as potentially significant in the simulation model were investigated in a further data analysis that is presented in chapter 11. A number of the predictions about phenomena resulting from the model could be shown to occur in the experimental data recorded, and none of them was clearly refuted. A number of other interesting possibilities for further *post hoc* data analysis and classification, partially stemming from the simulation model, partially from increased understanding of the data have not been pursued to their end. This is partially due to the fact that data analysis is time expensive. However, it also appeared necessary to find a cut-off point for the analysis of the measured data because the analysis presented pointed out in how far the experiment was sub-optimal or defective. Instead the insights gained so far should be used to design better, simpler and less ambiguous experiments to test the refined hypotheses resulting from this project (see following sections).

12.2 The Sensorimotor Basis of Present-Time Experience

The hypothesis tested in the experimental set-up was that training with sensory delays would lead to a decrease of performance in the post-test compared to the initially recorded performance in the pre-test, a hypothesis that was not confirmed by the data. Investigating the sensorimotor dynamics of behaviour *post hoc* and with the aid of ER simulation modelling gives reason to believe that this failure to confirm the experimental hypothesis was not because the ideas tested were faulty, but because the experimental set-up was faulty. The data supports the prediction generated by the ER simulation model that a negative after-effect in the form of micro-displacements occurred but was not registered by the coarse performance criterion.

As argued in chapter 3 section 3.6, this form of *post hoc* data analysis is more credible than an exhaustive data mining of the space of descriptive variables in search for one that matches the expected pattern sometimes associated with the concept of *post hoc* analysis. The simulation model generated a small number of very specific predictions about the kind of sensorimotor phenomena that can explain the experimental results as an artefact of the experimental set-up rather than a complete negation of the ideas underlying the hypothesis tested.

What can be inferred from these findings? In principle, the insights about the discrepancy between sensorimotor adaptation and task performance should make it possible to design a better experiment in which these variables concur. In simulation, using a fitness function that is spatially more exact had exactly this effect. Unfortunately, this modification is not feasible for the real experiment because the temporal sampling rate and the spatial resolution have fierce limits imposed by the fact that the experimental platform needs to work in real-time. But, even if these technical limitations could be mitigated, the thorough analysis of the participants' and the evolved agents' behaviour and its contemplation in the context of the general question of the sensorimotor basis of time cognition proposes a more profoundly different direction for further experimentation.

As a result of the experiments performed and earlier studies on adaptation to sensory delays (chapter 8, section 8.3), I want to introduce the classification of behavioural feedback loops into *reactive, reflex-like* and *anticipatory* (cf. Rohde & Di Paolo, 2007). Behavioural strategies that I call reactive are those in which the motor output is, at any point in time, contingent only on current sensory input. Phototaxis in a Braitenberg vehicle is the typical example of a reflex-like behaviour. As discussed in chapter 10, the strategies evolved in the model of the experimental

catch task are not reactive but instead reflex-like in the sense that the motor output associated with an action is only sensitive to stimulus onset, not to the moment to moment variation of stimulus magnitude. A third class of behavioural strategies is what I call 'anticipatory' strategies, in which the motor output is contingent on both moment-to-moment variation of inputs and the history/change of sensory inputs. These definitions are not fully formal (or even not at all) at this moment and aim more at capturing our intuitions about the differences between these kinds of strategies than at proving anything. Also, any behavioural strategy will probably be rather *more* or *less* reactive/reflex-like/anticipatory, rather than purely reactive/reflex-like/anticipatory.

In these different kinds of behaviour, sensory delays will play a different functional role. In purely reactive sensorimotor loops, a sensory delay implies that behaviour has to be slowed down. If the action at any point in time depends on the current sensory state, in a closed sensorimotor loop, this means that I have to wait in order to perceive the outcome of my previous action. A Braitenberg vehicle with long sensory latencies will turn past the light source, overshoot and eventually oscillate, just as the participants in Smith and Smith's (1962) outline-drawing task. This overshooting effect due to sensory delays in reactive sensorimotor loops is similar to the effect of an increase in inertia, such as in driving or canoeing, and can be compensated for by slowing down. To an extent, this was already recognised by Cunningham et al. (2001a). This led them to the conjecture that negative after-effects failed to occur in previous studies because they provided the possibility to compensate for the delay by slowing down.

In reflex-like behaviours such as those evolved in the artificial agents, in contrast, a delay corresponds to a fixed spatial offset that depends on the initial movement velocity. This completely different perturbation induced by a delay is what produces the systematic displacements observed in the data.

It is important to realise that in both kinds of behaviour, reactive and reflex-like, the perturbation effect induced by the prolonged sensory delay can be thus conceptualised as something which is not a delay (i.e., a displacement or an increase in inertia). Displacements and increases in inertia are perturbations we frequently suffer in our everyday lives and we are, therefore, very used to compensating for them. Furthermore, increases in mass/inertia do not tend to produce negative after-effects. Displacements produce spatial negative after-effects such as those observed in our experiment. I believe that these alternative conceptualisations are what hinder semi-permanent adaptation and the readjustment of experienced simultaneity - the delay is not experienced as a delay in the first place.

Therefore, Cunningham et al.'s (2001a) observation that "sensorimotor adaptation requires subjects to be exposed to the consequences of the discrepancy" is not fulfilled - subjects are exposed to the consequences of *some* discrepancy, but not of a discrepancy clearly identifiable as delay. It is not only necessary for the delay to impact on an established behavioural pattern, but also for it to impact *as a delay* on it, not as an increase in inertia or a displacement. While an increase in inertia makes you inherently slow, a delay still allows you to act fast - only delayed. Therefore, in order for a delay not to be conceptualised as an increase in inertia, it is necessary to include time pressure into the task, such that this difference between the two kinds of perturbation becomes significant. In a not fully conceptualised understanding of these matters resulting from Cunningham et al.'s study and hypothesis, I used high object velocities in my experiment. As

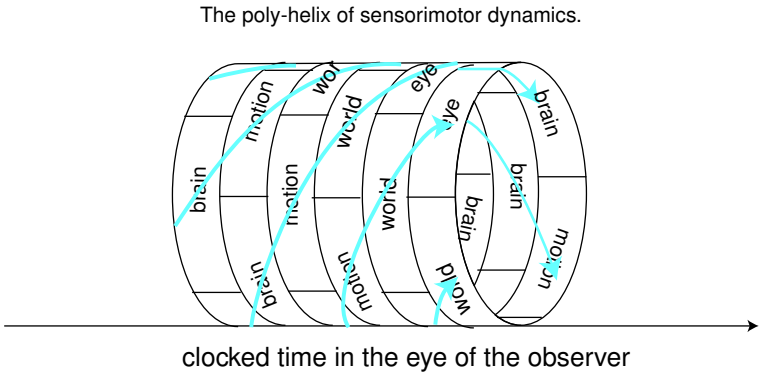

Figure 12.1: Illustration of ideas on temporal experience from the observer perspective and sensorimotor loops.

desired, no reactive solutions, in which subjects can slow down to compensate, were adopted by the subjects.

It turned out, however, that this is not enough. There is more than just time pressure to the experimental set-up used by Cunningham et al. Their visual task forces subjects to produce fast variable velocity motor sequences, in which a sensory delay can be neither understood as an increase in inertia, nor as a fixed displacement. The task they posed required an *anticipatory* strategy. This kind of behaviour is, however, only possible if the signal structure is rich enough to be systematically linked to the strategy over a longer period of time. There needs to be a cohesion between momentary signal structure, own movement possibilities, and future signal structure. The richness of the signal used in Cunningham et al.'s study was lost in my minimalist version of it, as it had not been assumed that this richness is necessary. From the failure of the experiment and, in particular, from understanding the sensorimotor dynamics of it, I now know that this assumption is false and why.

What do these findings and conceptual analyses teach us about the sensorimotor dynamics of time perception? How are sensorimotor latencies involved in the constitution of experienced presentness, pastness and futureness? I have some very tentative ideas that are not yet fully developed about the role of the absolute temporal length of a sensorimotor loop (i.e., the time it takes, from the observer perspective, for a stimulation to take effect in motion and, via the world, again on sensation) in constituting primitive past, present and future. These ideas relate to the analysis of the use of spatial and temporal language in the Aymaran vs. the English language and the role that knowledge and the possibility for agency plays in it (cf. chapter 8, section 8.2.3 and (Núñez & Sweetser, 2006)). What happens during a causal sensorimotor loop that is extended from the observer perspective is neither known to me, the agent, nor is it something that I can still act upon. Therefore, it is neither past (because the past is known and done and unchangeable) nor future (because the future is open to be changed by me or external forces). It is the present, that which is in its making to become the past once I know (through re-afferent sensation) that external forces do not interfere with my expectation of the outcome of my action.

At any moment in time, many sensorimotor loops are being realised. I tried to capture this idea in the diagram depicted in figure 12.1, in which subjective time, from the observer perspective,

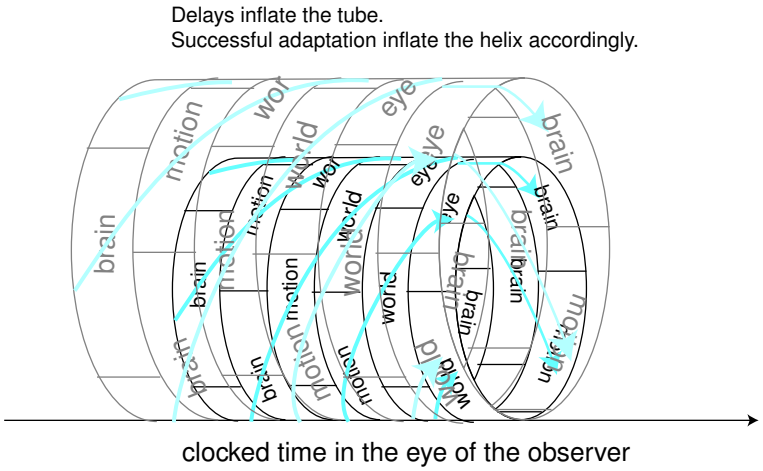

Figure 12.2: The tubular illustration of temporal experience from the observer perspective is inflated over adaptation to sensory delays.

takes the form of a tube. In this tube, change of variables in the eye of the observer forms the x-axis and the helices running around this tube are the causal sensation action loops, also in the eye of the observer.

These ideas are highly hypothetical and underdeveloped and possibly not well explained. However, they do account for both Libet's (2004) and Cunningham et al.'s (2001a) counterintuitive results on disruptions of experienced presentness: in Libet's experiments, the 500 ms that elapse between a peripheral stimulus and the build up of correlated cerebral activity, as well as the 500ms between the Readiness Potential and the onset of movement, form part of a sensorimotor loop, of a causal chain that is in its making and that cannot be changed through the subjects volition anymore. Therefore, these 500 ms in the eye of the observer, from the subject perspective, are neither future (changeable), nor past (confirmed truth). They are present, because they do not exist in any meaningful way temporally for the subject itself, unless, through the use of technology, I bring them into existence by cutting short the sensorimotor latencies. Similarly, in Cunningham et al.'s experiment by imposing a delay, the sensorimotor loop was stretched such that the extra 200ms became a meaningless time span, one in which subjects could neither act nor react, and therefore were banned from temporal experience. This corresponds to inflating the tube of temporal experience depicted in figure 12.2. If the delay is removed, the tube is shrunk and the subject is suddenly afforded an extra 200 ms to become active - which, initially is experienced as an inversion of the temporal order associated with causal chains.

These ideas in progress will have to be worked out and explicated, the preceding three paragraphs do not intend to provide a full-blown theory of the sensorimotor basis of time cognition. They only express some thoughts in progress that will surely influence the design of future experiments and, through these experimental results and simulation modelling, hopefully become explicated and illustrated.

12.3 Future Research

The experiment on adaptation to sensory delays and their role in the constitution of present-time experience had tested the hypothesis that time pressure is what is necessary to yield a semi-permanent adaptation in the strong sense outlined in chapter 8 section 8.3. This hypothesis had not been confirmed. Simulation modelling and *post hoc* data analysis of the sensorimotor data recorded gives reason to believe that the hypothesis was not wrong because there is no relation between time pressure and this strong form of adaptation, but because it may not be the only factor necessary. The theoretical and empirical insights from this project will inform future experiments on the same matter, and these experiments will employ, again, the minimalist modelling and experimental framework developed.

However, the insights gained are not going to lead to a modified version of the previous experiment. Much rather, they will lead to the conduction of an entirely new set of experiments that are, in a way, more complex, to account to the more sophisticated new hypothesis, and, in a way, more simple, to constrain the strategy space afforded by the experiments in order to avoid the heterogeneity in behaviour, perturbation and adaptation that marked the data gathered in the experiment presented in chapter 9.

The insights about different kinds of sensorimotor loops (reactive, reflex-like and anticipatory) suggest that, apart from time pressure, a structured environment/signal is necessary: the former to suppress reactive compensation and conceptualisation of the delay as an increase in inertia, and the latter to make anticipatory behaviour possible. This new hypothesis can be tested using the interdisciplinary approach proposed. There are, indeed, plans for collaboration with experimental psychophysics groups in order to pursue these matters beyond my doctorate studies.

Ideas for experiments to test the refined hypotheses include a replication of Cunningham et al.'s (2001a) study with increased emphasis on sensorimotor dynamics. Even though the visual component in the original task makes a full dynamical analysis and ER modelling difficult, it would be worthwhile to take a look at such data as to whether it supports the ideas presented here or not. Additional control conditions with delay but without time-pressure/predictability could give additional evidence for the hypothesis here developed.

Other ideas and experimental paradigms have been thought about. For instance, an extension of the vestibular feedback condition tested by the same group (Cunningham et al., 2001c) would be interesting, because the sensory dimension is simpler and easier to model. PS techniques could be used in the same set-up, comparing 'artificial' vestibular sense and natural vestibular sense. I also have the idea to develop a 'temporal Necker cube', i.e., a set-up in which the meaning of a situation is ambiguous as concerns its temporal interpretation. Identifying conditions that provoke one interpretation or the other would give powerful evidence for the theories proposed.

This is not the time or the place to spell out future experimental set-ups in detail. One of the main lessons from this project is that careful planning is essential and I will not make the mistake to rush experimental planning. Experts will have to be approached with the ideas just outlined for consultancy on how to devise sound and non-ambiguous conditions with clear fallback positions.

Another point to mention is that the institute in which Cunningham et al. conducted the mentioned studies (i.e., the Max-Planck-Institute for Biological Cybernetics) have recently found a different paradigm that appears to produce the interesting semi-permanent adaptation to delays

with changes in experienced simultaneity: participants waving their hand under a TV screen appear to adapt to delayed visual feedback, if the visual image on the TV screen covering their hand is a delayed image of their hand (Ernst and Frissen, unpublished work; personal communication Sept. 2007). These new findings will have to be incorporated and evaluated with respect to the ideas presented and may indicate further possibilities for minimal experimental set-ups.

The avenues for future research are open-ended and it is my strong personal desire to pursue the ideas presented in this second half of the dissertation beyond this degree, using and improving the interdisciplinary approach proposed.

Chapter 13

Conclusion

This last chapter summarises and evaluates the results presented in chapters 4-12 with respect to the methodological research question outlined in chapters 2 and 3, i.e.: Can ER simulation models serve as a scientific tool in an enactive approach to the scientific study of human cognition and behaviour, and if yes, how?

Section 13.1 summarises the motivation, research question and results, before section 13.2 evaluates them. The conclusion section 13.3 locates the contribution made by this dissertation in the larger research context and points to its applicability and extendibility in future research.

13.1 Summary

In trying to move beyond the paradigmatic struggle in Cognitive Science, this dissertation promotes and methodologically advances the enactive approach to human cognition and behaviour. For historical reasons, the metaphor of cognition as computation is closely tied to the idea of the interdisciplinary and scientific study of mind and cognition, in particular if it involves computer modelling. Over the past decades, however, the computational metaphor turned out to be empirically limited and conceptually harmful. The enactive approach rejects this metaphor in favour of an embodied, situated, dynamical and constructivist perspective that focuses on autonomous dynamics on several emergent levels of biological organisation, on experience and on the genuine meaningfulness of mind and mindful behaviour. In chapter 2, I introduced this debate, argued for the impossibility to compromise on the level of paradigms and identified the challenges that lie in the future of the enactive approach. I identified processes of very abstract, symbolic and high level cognition as representationalist strongholds that pose the biggest challenge to the enactive paradigm to demonstrate its explanatory potential. At the end of chapter 2, I outline the central methodological question addressed in the present dissertation, i.e., how ER simulation modelling as a technique for enactive Cognitive Science can be used in order to elucidate aspects of human cognition and behaviour, and particularly how it can contribute to invading representationalist strongholds.

From there, the repertoire of methods used in this dissertation (ER simulation modelling, CTRNN controllers, DST analysis and PS experiments) was introduced in chapter 3. However, as

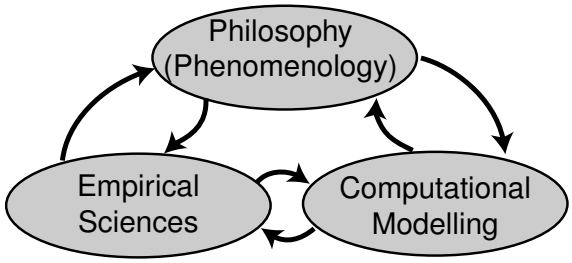

Figure 13.1: Illustration of the interdisciplinary enactive approach proposed and tested.

this dissertation has a methodological emphasis, chapter 3 also saw an extensive debate on the issues of the role of the scientist as an observer in constructivist approaches, on the scientific value of ALife simulation models and on the possibility of scientifically studying experience by combining first and third person methods in a non-reductive fashion. Crucially, this chapter also developed the interdisciplinary methodological framework put to use in the later parts of this dissertation, i.e., the application of ER simulation modelling to minimalist experiential and experimental research on perception and sensorimotor adaptation (PS research). It concludes that this kind of research is only truly interdisciplinary if modelling, experimentation and personal experience are considered at the same time and to an equal extent, mutually informing each other (see figure 13.1). The results presented in the subsequent chapters aim at highlighting individual parts of this diagram, building up to the study on adaptation to sensory delays for which I not only provided the model, but also conducted the experimental study, working in all the disciplines depicted in figure 13.1 and implementing all six mutual methodological links.

Chapter 4 presented a model of directional reaching in an idealised human arm to investigate the principle of linear synergies in motor organisation. This model shows that imposing this kind of constraint on the motor system enhances evolvability. This benefit is not just due to the smaller search space: reducing the task to two dimensions means an equal decrease of parameter space but has the opposite effect on evolvability. Concerning the methodological theme of the dissertation, this study successfully implemented the links between simulation modelling and the experimental sciences, as discussed in the following section 13.2.

The simulation model of value system architectures presented in chapter 5 investigated a research question of a much more abstract and philosophical nature, i.e., it illustrates logical problems with a certain type of neural or cognitive architecture and points out the implicitly held modelling assumptions underlying such approaches. The model criticises *a priori* semantics of homuncular modules in an embodied context as a proposed solution to problems encountered with a fully homuncular approach. It points out how such homuncular modules, if they are postulated rather than explained, are unlikely to explain cognition or adaptivity as a general phenomenon. This model was a demonstration of the mutual methodological links between philosophical theory building and simulation modelling.

The following two simulation models on perceptual crossing in a one-dimensional (chapter 6) and a two-dimensional (chapter 7) simulated environment applied ER modelling to experiments in PS. The simulation models contribute towards the understanding and interpretation of the experiments on a number of different levels of description, from generating concrete hypotheses about

variables involved in perceptual distinction and morphological aspects of observed behaviour to providing abstract proofs of concept about dynamical principles at work and implicit premises held by experimenters or subjects. As this approach already worked on the interdisciplinary link between experimental and experiential-philosophical methods, these models succeeded at implementing all four methodological links between simulation modelling and the other disciplines in figure 13.1 at a time.

Finally, the interdisciplinary project on experienced simultaneity and adaptation to sensory delays presented in chapters 8 - 12 aimed to identify the minimal conditions under which a semi-permanent adaptation to sensory delays that involves a distortion of experienced simultaneity occurs. In this project, I did not only provide the model and the mutual links to experiential-conceptual debate and experimental practice respectively, I tried to engage in all three to equal extents, and to thereby realise the entire interdisciplinary approach depicted in figure 13.1 in person, though with the help of experienced collaborators. This project is, in its very nature, more complex than the previous simulation models and therefore consumes large parts of this dissertation (and the time elapsed to conduct the work presented). The original hypothesis tested in the experimental study was not confirmed. However, the interdisciplinary analysis of the sensorimotor dynamics of behaviour and a simplified ER model of the task still produced valuable results and revised hypotheses for further experimentation and conceptual advances towards a theory of the sensorimotor basis of time experience. The results in terms of acid testing the methodological framework developed in chapter 3, though mainly positive, are mixed; its merits and demerits are evaluated in detail in the following section 13.2.

13.2 Evolutionary Robotics Simulation Models in the Study of Human Behaviour and Cognition

The conclusion from the work presented in the current dissertation with respect to the overarching methodological research question, i.e., in how far ER simulation models can be used as a tool in an enactive science of human adult cognition, is overall very positive. Chapter 3 identifies the theoretical possibilities of minimal simulation models to contribute to scientific explanation. The subsequent result chapters provide case studies that, by addressing different research questions, exemplify these kinds of possible contributions. In this section, I want to evaluate three issues more profoundly: the question of the recognition and incorporation of simulation results in the targeted fields (section 13.2.1), the question of advancing the enactive paradigm by conquering representationalist strongholds (section 13.2.2) and a more detailed critique of the interdisciplinary framework proposed in chapter 3 that was applied in the project on adaptation to delays in chapters 8-12 (section 13.2.3).

13.2.1 Feeding Results Back to the Relevant Scientific Community

With simulation modelling, a potential danger is that its results get lost in a nexus. It is inspired by a real biological phenomenon, models this phenomenon, thereby proves important principles for synthetic approaches, but also important lessons for the original experimental research domain, but does not succeed in feeding these results back into either of the relevant communities or receiving the deserved attention and acknowledgement. Concerning the integration of simulation results I

presented into the research context they address, I can be more than happy with the impact they had in the relevant scientific communities.

Both groups working on motor synergies that had inspired the model presented in chapter 4 were very positive, suggested extension of the work and cited my work in a high impact publication (Shemmell et al., 2007). Had I pursued this strand of research, I am sure that future collaboration and dialogue with the groups involved would have been possible.

Concerning the critique of value system architectures presented in chapter 5, it has not been as successful in impacting on the relevant communities. My personal belief is that this is not because the model lacks merit, but because it has not been well presented to the relevant communities and because it has focused on decomposition rather than on identifying possibilities. Extending, re-interpreting and publishing the results is a task that remains in order to confirm this assessment of mine.

The models of perceptual crossing in a one-dimensional and a two-dimensional simulated environment have been well received by the group (GSP) who conducted the original study and cited as a relevant contribution (Auvray et al., 2008). It is worthwhile pointing out that this modelling work was conducted before, during and after my placement in the GSP, where I realised the project on the adaptation to sensory delays such that communication of the results and arguing their significance was immensely facilitated. We could publish our results and the ER simulation approach taken in a domain-internal (i.e., a psychological) journal (Di Paolo et al., 2008b).

For the project on adaptation to delays whose results are not published, it remains to be seen how it is received in the rather conservative psychophysics and perception research community. This is a problem that concerns the minimalist experimental and experiential research in sensorimotor behaviour and PS in general. In presenting the results, an additional problem is that technical problems were associated with the realisation of the experiment and that, in order to provide a sound scientific contribution rather than just a methodological exploration, the experiments may have to be repeated. Since the ideas underlying the study have been revised, it is unlikely that this repetition will be done. I believe, however, that if valid results were thus obtained, they would be recognised and acknowledged in the relevant community, if they are well argued and presented. These are, however, tasks that remain.

Overall, I can conclude that I have received surprisingly enthusiastic and positive reactions from the scientific communities targeted. I do, however, believe that my considerable efforts in contacting these groups, arguing and communicating the principal merits and exact contributions of my model and general dissemination efforts were necessary to obtain these kinds of reactions. I advise other researchers working with minimal simulation modelling approaches to pursue the same path if they do not want to see their work disappear in the nexus described.

13.2.2 Invading Representationalist Strongholds

The ideas of high and low levels of cognition are very different between a representationalist and an enactive perspective on cognition. In a representationalist view, high level cognition is the kind of symbol manipulation performed in the most decoupled and homuncular modules that are furthest away from the sensory and motor periphery. In an enactive perspective, it is not fully clear how high or low level cognition should be defined other than in phylogenetic advances and new

forms of value generation on new and more abstract levels of autonomous self sustaining dynamics (see chapter 2 and chapter 5 section 5.1). It is important to realise that for the latter concept of high level cognition, peripheral systems of the organism are in no way inferior or less important than high level brain areas. This is the whole idea about embodiment and the closed sensorimotor loop. Therefore, there is no *a priori* congruence between what is considered high and low level cognition from these two perspectives.

Concerning the model of motor synergies (chapter 4), however, it is probably considered more low level in both perspectives. From the representationalist perspective because it is concerned with the realisation of motion, not with motor planning or reasoning. From the enactive perspective, it is low level because the processes described and investigated do not form part of the experiential world and are very closely tied to the mechanical domain of the organism, even if they have interesting implications for the functional domain. This is not to say that the study of principles in motor control is irrelevant or boring, on the contrary. It is, however, the reason why I eventually abandoned this rewarding area of research in favour of tackling questions of higher levels of cognition.

The model of value system architectures (chapter 5), on the contrary, dives straight into the question of fundamental neural organisation and its functional role in constituting cognition and adaptivity as a general phenomenon. From the representationalist perspective, it is probably impossible to go any more high level than that. For that very reason, however, it is a largely conceptual model. It criticises the idea of localised *a priori* semantics in popular hybrid and semi-homuncular approaches. However, it has nothing to put in its place. This failure to provide concrete and empirically testable ideas seems logical in the light of the enactive belief that cognition should not be studied in neural modules but much rather as a global and irreducible phenomenon. Targeting cognition as a general phenomenon, simple simulation models can only be interesting for science from within a representationalist perspective that assumes that functional modules will ultimately add up in a linear fashion. Stepping outside this approach and the assumption of local reducibility of function, a simulation model is confined to fuelling conceptual, high level debate like the model I provided. Again, though I believe these contributions to be perfectly valuable and honourable, concerning my own personal research interests, I felt the urge to contribute something more meaty and tangible to Cognitive Science. I wanted to contribute more immediately to the concrete scientific practice of explaining mind.

This aspiration of mine led me to seek the links with the GSP and their minimalist PS research (chapters 6-12). They scientifically investigate questions of high level cognition (perceptual experience) from an enactive and methodologically minimalist perspective. A more coincidental benefit of applying ER modelling to this approach is that, in most of the group's studies, they use simulated environments that are as simple as those used for ER modelling. Using the same simulation in both allows to draw stronger analogies between the two, which allowed me to generate a number of concrete quantitative predictions about the behaviour obtained in the spirit of a more descriptive modelling approach. This approach allowed me to combine the aspiration of addressing questions of genuine cognitive interest and contributing directly to scientific practice with my models.

The experiments I modelled (or conducted) addressed the problems of minimal social interac-

tion and perceived intentionality (perceptual crossing) and the sensorimotor basis of present time experience. Both of these are, to my understanding, fascinating problems of the most abstract and high level cognitive nature that can, in this form, only be found in humans and that are amongst the most fundamental dimensions of the human mind. It is maybe not surprising that this perception of the sophistication of the *explanandum* is not necessarily shared by the representationalist community. Talking to representationalist peers about this kind of research, they were only moderately impressed. Still trapped in a representationalist conception of mind, brain and cognition, it seemed impossible for them to accept that these kinds of minimal sensorimotor experiments can do more than just explain the part played by low level processes of automatic time detection, agency detection or coordination in building up an internal representation of the world for cognition to work on and to serve the actual processes of symbolic cognition. This divergence in credit from both communities, is, in my opinion, not a shortcoming, but a possibility. By generating scientific explanations of the perceptual experience of intentionality, time or space, we can invade these representationalist strongholds and, ultimately, make the ideological building collapse by bringing it into conflict with the data we provide.

13.2.3 Evaluating the Interdisciplinary Framework Proposed

The research conducted within the scope of my doctorate saw an increase in methodological adventurousness. The kind of scientific contribution that the models of motor synergies and value system architectures provided had been previously argued and analysed (e.g., Harvey et al., 2005; Di Paolo et al., 2000; Beer, 1996). However, the application of ER modelling to PS research in a close match between experiment and model was a methodological novelty. The models of perceptual crossing could, though novel in the kinds of results they generate, be expected to follow similar rules and principles as other ER simulation models. In contrast, the format and characteristics of the proposed interdisciplinary framework, in which three methodological domains - the experimental, the experiential and synthetic modelling - are exercised to equal extent and by the same research, with mutual influences between all three, were of largely exploratory nature. In this section, I want to evaluate this proposal and its implementation in the study of experienced simultaneity and adaptation to sensory delays, paying special attention to the questions of the experiential dimension of the work and the organisation of scientific activity in the mentioned domains.

As concerns the equal proportion of all three classes of approach, I discuss already in chapter 3 section 3.5 that the experiential dimension has been methodologically underdeveloped and has, as a consequence, not received the attention it deserves in practice. This had been recognised as a failure of mine, not of the interdisciplinary framework developed. In section 3.5 I analyse the possibilities to draw connections between third, first and second person methods and propose to include the perceptual measures used in classical psychophysics as alternative first/second person methods into the neurophenomenological approach outlined by Varela (1996). In future research, this direction will definitely be pursued in order to come closer to the idea of giving equal weight to the three methodological domains.

Another aspect that had not been fully developed in chapter 3 section 3.6 is in how far it is necessary or beneficial that the methods identified are employed by the same person and how the

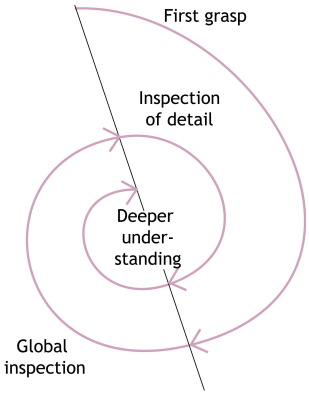

Figure 13.2: Illustration of the hermeneutic circle of understanding.

use of such methods is temporally organised. The implicitly held assumption was that performing both the experiment and implementing the model in person would lead to a much closer interaction between the two, such that modelling and scientific practice would constantly mutually inform each other and keep growing alongside one another. In practice, however, I found that there were clear phases of working conceptually, experimentally or with the model - I did *not* actually develop them alongside one another.

Is this a property of the outlined approach in general or of my way of going about it? As already argued in (Rohde & Di Paolo, 2007), I believe that starting the modelling work earlier could have saved me from some of the frustration associated with the unconfirmed main hypothesis. However, I believe that the modelling work and the experimental work require two different perspectives and that they can be exercised at the same time only to a certain extent. This resonates with the classical idea of the hermeneutic circle of understanding of a text described by Gadamer (1994, recent edition; German original published in 1960), in which understanding is advanced by alternating phases of closure and prejudice, from the global perspective, and phases of thorough investigation of detail, in which our ideas are open to change (see figure 13.2). The analogy is not fully valid as both simulation and experimental planning/measurement are relevant in both the global and local phases of understanding. However, I think a similar diagram can be used in order to illustrate how a phase of modelling can aid experimental design, be pushed into the background during piloting, return to the foreground for elaboration of the set-up, become irrelevant during conduction of the experiment, but later aid interpretation of the results, etc.

In so far, the benefits of performing all tasks in the interdisciplinary framework (figure 13.1) in person (as opposed to contributing with simulation modelling to existing experimental research, as in the modelling of perceptual crossing), are only of a quantitative nature, not of a qualitative nature, as I had previously assumed. By that, I mean that conducting both the experimental research and the modelling assures that the mutual links between these disciplines work smoothly and the relevance of one for the other is taken into account. However, a close collaboration and working communication between researchers working in both areas is, in principle, equally powerful.

In conclusion, from the experience of putting the interdisciplinary framework proposed in this dissertation (figure 13.1) to work, I am more than ever convinced of its merits. However, there are still substantial parts of it that have to be developed and explicated. In particular, the

experiential dimension has to be methodologically advanced and practically incorporated, and the organisational practice of working within the participating disciplines and interdisciplinarily in terms of time, space, management, communication etc. have to be explicated and developed. In chapter 3 section 3.6, I lamented that the reductionist traditional computationalist approach is not actually interdisciplinary but rather multidisciplinary in that research on the different levels that are reduced to one another can be conducted more or less independently. This is not the same in the enactive approach, which is a genuinely interdisciplinary framework. This does, however, make things much more difficult - a lot of work lies ahead of us to discover, develop and explicate the inventory of the enactive toolbox.

13.3 Straight Ahead

As outlined in the introductory chapter 1, the research presented in this dissertation can be seen as an episode in a personal journey driven by curiosity and nagging questions about how the mind works and how it relates to the world. This journey is not finished with the submission of this dissertation. I have addressed a number of different research areas and questions with ER simulation modelling, some of which I have brought to a temporary conclusion. However, the methodological possibilities identified, and particularly the combination of ER simulation modelling with minimal experiments on sensorimotor behaviour such as the GSP's PS approach or some types of psychophysics research, which I see as the main research contribution of this dissertation, clearly calls for further application and development. This and the previous chapter, which evaluated the interdisciplinary study on adaptation to sensory delays, conclude with more open questions than with answers - luckily. Because it means that interesting times lie ahead of me.

Bibliography

Allen, J. (1984). Towards a general theory of action and time. *Artificial Intelligence*, *23*, 123–154.

Amedi, A., Stern, W., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*, 687 – 689.

Arbib, M. (1981). Perceptual structures and distributed motor control. In Brooks, V. (Ed.), *Handbook of Physiology*, Vol. II, Motor Control, Part 1, pp. 1449–1480. American Physiological Society. Section 2: The Nervous System.

Ashby, W. (1954). *Design for a Brain*. Chapman and Hall Ltd., London.

Auvray, M., Lenay, C., & Stewart, J. (2008). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*. (Forthcoming).

Bach-y Rita, P., Collins, C., Sauders, F., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, *221*, 963–964.

Bach-y Rita, P., Tyler, M., & Kaczmarek, K. (2003). Seeing with the brain. *Int. J. Human-Computer Interaction*, *15*, 285–295.

Baird, J., & Noma, E. (1978). *Fundamentals of Scaling and Psychophysics*. John Wiley & Sons, New York. Wiley Series in Behaivor.

Barandiaran, X. (2008). Mental Life: Conceptual models and sythetic methodologies for a post-cognitivist psychology. In Wallace, B. (Ed.), *The World, the Mind and the Body: Psychology after cognitivism*. Imprint Academic. (In press).

Bedford, F. (1993). Perceptual learning. In Medin, D. (Ed.), *The Psychology of Learning and Motivation*, Vol. 30, pp. 1–60. Elsevier.

Beer, R. (1995). On the dynamics of small Continuous-Time Recurrent Neural Networks. *Adaptive Behavior*, *3*(4), 469–509.

Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In Maes, P., Mataric, M., Meyer, J., Pollack, J., & Wilson, S. (Eds.), *From Animals to Animats 4*, pp. 421–429. MIT press.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*, 91–99.

Beer, R. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, *4*(11), 209–243.

Beer, R. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, *10*, 309–326.

Beer, R. (2006). Parameter space structure of Continuous-Time Recurrent Neural Networks. *Neural Computation*, *18*, 3009–3051.

Bernstein, N. (1967). *The Coordination and Regulation of Movements*. Pergamon, Oxford. Russian original published in 1935.

Bertschinger, N., Olbrich, E., Ay, N., & Jost, J. (2008). Autonomy: An information theoretic perspective. *BioSystems*, *91*(2), 331–45. Special issue on modelling autonomy.

Bitbol, M. (2001). Non-representationnalist theories of cognition and quantum mechanics. *SATS (Nordic journal of philosophy)*, *2*, 37–61.

Breese, B. (1909). Binocular rivalry. *Psychological Review*, *16*, 410–415.

Brooks, R. (1995). Intelligence without reason. In Steels, L., & Brooks, R. (Eds.), *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*. Lawrence Erlbaum, Hillsdale, NJ.

Cantwell-Smith, B. (1996). *On the Origin of Objects*. MIT Press, Cambridge MA.

Chalmers, D. (1995). Facing up to the problem of consicousness. *Journal of Consciousness Studies*, *2*, 200–220.

Chrisley, R. (2003). Embodied Artificial Intelligence. *Artificial Intelligence*, *149*, 131–150.

Churchland, P. M., & Churchland, P. S. (1998). *On the Contrary: Critical Essays 1987-1997*. MIT Press, Cambridge MA.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press, Cambridge MA.

Clark, A. (1998). Time and mind. *The Journal of Philosophy*, *95*(7), 354–376.

Clark, A., & Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*, *7*, 5–16.

Cliff, D. (1991). Computational Neuroethology: A provisional manifesto. In Meyer, J., & Wilson, S. (Eds.), *Proc 1st Int. Conf. on Simulation of Adaptive Behaviour: From Animals to Animats*, pp. 29–39 Cambridge MA. MIT Press.

Cunningham, D., Billock, V., & Tsou, B. (2001a). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science*, *12*, 532–535.

Cunningham, D., Chatziastros, A., von der Heyde, M., & Bülthoff, H. (2001b). Driving in the future: Temporal visuomotor adaptation and generalization. *Journal of Vision*, *1*(2), 88–98.

Cunningham, D., Kreher, B., von der Heyde, M., & Bülthoff, H. (2001c). Do cause and effect need to be temporally continuous? Learning to compensate for delayed vestibular feedback. Abstract. *Journal of Vision* 1(3): 135a.

De Jaegher, H. (2007). *Social Interaction Rhythm and Participatory Sense-Making. An Embodied, Interactional Approach to Social Understanding, with Implications for Autism*. Ph.D. thesis, Department of Informatics.

Dennett, D. (1985). *Elbow Room: The Varieties of Free Will Worth Wanting*. Clarendon Press.

Dennett, D. (1989). *The intentional stance*. MIT Press, Cambridge MA.

Di Paolo, E. (2000a). Behavioral coordination, structural congruence and entrainment in acoustically coupled agents. *Adaptive Behavior*, *8*, 27–47.

Di Paolo, E. (2000b). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior* Cambridge MA. MIT Press.

Di Paolo, E. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In Murase, K., & Asakura, T. (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*, pp. 19–42. Advanced Knowledge International, Adelaide, Australia.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452.

Di Paolo, E., & Iizuka, H. (2008). How (not) to model autonomous behaviour. *BioSystems*, *91*, 409–423. Special issue on modelling autonomy.

Di Paolo, E., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In *Artificial Life VII: The Seventh International Conference on the Simulation and Synthesis of Living Systems, Reed College, Portland, Oregon, USA, 1-6 August (Proceedings)*.

Di Paolo, E., Rohde, M., & De Jaegher, H. (2008a). Horizons for the enactive mind: Values, social interaction, and play. In Stewart, J. Gapenne, O., & Di Paolo, E. (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*. MIT Press, Cambridge, MA. (in press).

Di Paolo, E., Rohde, M., & Iizuka, H. (2008b). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*. Special Issue on Dynamical Systems approaches (in press).

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*, 495–506.

Eagleman, D., & Sejnowski, T. (2002). Untangling spatial from temporal illusions. *Trends in Neurosciences*, *25*, 293.

Edelman, G. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books, New York.

Edelman, G. (1989). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books, New York.

Edelman, G. (2003). Naturalizing consciousness: A theoretical framework. *Proc Natl Acad Sci USA*, *100*, 5520–5524.

Ehrenstein, W., & Ehrenstein, A. (1999). Psychophysical methods. In Windhorst, U., & Johansson, H. (Eds.), *Modern techniques in neuroscience research*, pp. 1211–1241. Springer, Berlin, Heidelberg.

Elman, J. (1998). Connectionism, Artificial Life, and Dynamical Systems: New approaches to old questions. In Bechtel, W., & Graham, G. (Eds.), *A Companion to Cognitive Science*. Basil Blackwood, Oxford.

Evans, V. (2004). How we conceptualise time: Language, meaning and temporal cognition. *Essays in Arts and Sciences*, *XXXIII*, 13–44. Issue Theme: Time.

Eysenck, M., & Keane, M. (2000). *Cognitive Psychology: A student's handbook* (Fourth edition). Psychology Press, Hove.

Fechner, G. (1966). *Elements of Psychophysics. Volume I.* Holt, Rinehartand Winston, Inc., New York. Translated by H. E. Adler. Edited by D.H. Howes and E. G. Boring. German original published in 1860.

Ferell, W. (1965). Remote manipulation with transmission delay. *IEEE Trans, hum. Factors Elect.*, *HFE-6*, 24–32.

Fodor, J. (2000). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*. MIT Press, Cambridge MA.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71. Special issue on Connections and Symbols, edited by S. Pinker and J. Mehler.

Fröse, T. (2007). On the role of AI in the ongoing paradigm shift within the Cognitive Sciences. In Lungarella, M. (Ed.), *50 Years of AI*, pp. 63–75. Springer.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*, 773 – 778.

Gadamer, H.-G. (1994). *Truth and Method*. Continuum, New York. Translated by J. Weinsheimer and D.G. Marshall. German original published in 1960.

Gapenne, O., Rovira, K., Ali Ammar, A., & Lenay, C. (2003). Tactos: Special computer interface for the reading and writing of 2D forms in blind people. In Stephanidis, C. (Ed.), *Universal Access in HCI: Inclusive Design in the Information Society*, pp. 1270–1274. Lawrence Erlbaum Associates, London.

Gepshtein, S., & Kubovy, M. (2007). The lawful perception of apparent motion. *Journal of Vision*, *7*, 1–15.

Gibson, J. (1982). The problem of temporal order in stimulation and perception. In Reed, E., & Jones, R. (Eds.), *Reasons for Realism: Selected Essays of James J. Gibson*, pp. 171–179. Lawrence Erlbaum, Hillsdale NJ. Originally published in the *Journal of Psychology*, 1966, 62, 141-149.

Gottlieb, G., Song, Q., Almeida, G., Hong, D., & Corcos, D. (1997). Directional control of planar human arm movement. *Journal of Neurophysiology*, *78*, 2985–2998.

Grossberg, S., & Paine, R. (2000). A neural model of corticocerebellar interactions during attentive imitation and predictive learning of sequential handwriting movements. *Neural Networks*, *13*, 999–1046.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Harvey, I. (1996). Untimed and misrepresented: Connectionism and the computer metaphor. *AISB Quarterly*, *96*, 20–27.

Harvey, I., Di Paolo, E., Wood, R., Quinn, M., & Tuci, E. A. (2005). Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, *11*(1-2), 79–98.

Haugeland, J. (1981). Semantic engines: An introduction to mind design. In Haugeland, J. (Ed.), *Mind design: Philosophy - Psychology - Artificial Intelligence*. MIT Press, Cambridge MA.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, MA.

Hebb, D. (1949). *The Organization of Behavior. A Neuropsychological Theory*. John Wiley & sons inc., New York.

Heidegger, M. (1963). *Sein und Zeit* (Tenth unmodified edition). Max Niemeyer Verlag, Tübingen. Original published in 1927.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell, Oxford.

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213*, 84–94.

Held, R., Efstathiou, A., & Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology*, *72*, 887–891.

Herzog, M. (2007). Spatial processing and visual backward masking. *Advances in Cognitive Psychology*, *3*, 85–92.

Hinton, G., & Nowlan, S. (1987). How learning can guide evolution. *Complex Systems*, *1*, 495–502.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

Holland, O. (2002). Grey Walter: The imitator of life. In Damper, R., & Cliff, D. (Eds.), *Biologically-Inspired Robotics: The Legacy of W. Grey Walter. Proceedings of the EPSRC/BBSRC International Workshop WGW-02*, pp. 32–48.

Hurlburt, R., & Schwitzgebel, E. (2007). *Describing Inner Experience? Proponent Meets Skeptic*. MIT Press, Cambridge MA.

Hurley, S. (1998). *Consciousness in Action*. Harvard University Press, London.

Hurley, S., & Noë, A. (2003). Neural plasticity and consciousness. *Biology and Philosophy*, *18*, 131–168.

Iizuka, H., & Di Paolo, E. (2007). Minimal agency detection of embodied agents. In Almeida e Costa, F., Rocha, L., Costa, E., Harvey, I., & Coutinho, A. (Eds.), *Proceedings of the 9th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, pp. 485–494 Berlin, Heidelberg. Springer.

Iizuka, H., & Ikegami, T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, *10*, 361–378.

Ikegami, T., & Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *BioSystems*, *91*, 388–400. Special issue on modelling autonomy.

Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Izquierdo-Torres, E., & Harvey, I. (2007). Hebbian learning using fixed weight evolved dynamical 'neural' networks. In Abbass, H., Bedau, M., Nolfi, S., & Wiles, J. (Eds.), *Proceedings of the First IEEE Symposium on Artificial Life*, Series on Computational Intelligence, pp. 394–401 Honolulu, Hawaii. IEEE.

James, W. (1890). *The Principles of Psychology*, Vol. 1. Chapter 15: The Perception of Time. Retrieved: 14.03.2008 from 'Classics in the History of Psychology. An internet resource developed by Christopher D. Green. URL: http://psychclassics.asu.edu/James/Principles/prin15.htm.

Jonas, H. (1966). *The phenomenon of life: Towards a philosophical biology*. Northwestern University Press, Evanston, IL.

Kandel, E., Schwartz, J., & Jessel, T. (Eds.). (2000). *Principles of Neural Science* (Fourth edition). McGraw-Hill, New York.

Kant, I. (1974). *Kritik der reinen Vernunft*, Vol. 1 of *Wissenschaft, die drei Kritiken*. Suhrkamp, Frankfurt a. M. Edited by W. Weischedel. German original published in 1787.

Kelso, S. (Ed.). (1982). *Human Motor Behavior: An Introduction*. Lawrence Erlbaum, Hillsdale, NJ.

Kirsh, D. (1991). Today the earwig, tomorrow man?. *Artificial Intelligence*, *47*, 161–184.

Kohler, I. (1962). Experiments with goggles. *Scientific American*, *206*, 62–72.

Krichmar, J., & Edelman, G. (2002). Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cereb. Cortex*, *12*, 818–30.

Kurthen, M. (1994). *Hermeneutische Kognitionswissenschaft. Die Krise der Orthodoxie*. DJRE Verlag, Bonn.

Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. University of Chicago Press. With an afterword.

Lakoff, G., & Núñez, R. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books, New York.

Langton, C. (Ed.). (1997). *Artificial Life: An overview* (First MIT Press paperback edition). MIT Press, Cambridge MA.

Le Van Quyen, M., & Petitmengin, C. (2002). Neuronal dynamics and conscious experience: An example of reciprocal causation before epileptic seizures. *Phenomenology and the Cognitive Sciences*, *1*, 169–180.

Lenay, C. (2003). Ignorance et suppléance : La question de l'espace. HDR. Université de Technologie de Compiègne.

Lenay, C., Gapenne, O., Hanneton, S., Marque, C., & Genouëlle, C. (2003). Sensory Substitution: Limits and perspectives. In Hatwell, Y., Streri, A., & Gentaz, E. (Eds.), *Touching for Knowing*, pp. 275–292. Benjamins Publishers, Amsterdam. Chapter 16, English translation.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354–61.

Libet, B. (2004). *Mind time. The temporal factor in consciousness*. Perspectives in Cognitive Neuroscience. Harvard University Press, Cambridge MA and London. Edited by S. Kosslyn.

Macho, T. (Ed.). (1996). *Wittgenstein*. Philosophie jetzt! Edited by Peter Sloterdijk. Eugen Diederichs Verlag, München.

Mann, T. (1985). *Buddenbrooks: Verfall einer Familie*. Fischer Taschenbuch Verlag, Frankfurt a. Main. Copyright 1922 - original published in 1901.

Markram, H. (2006). The blue brain project. *Nat Rev Neurosci*, *7*, 153–160.

Maturana, H. (1978). Kognition. In Hejl, P., Köck, P., & Roth, G. (Eds.), *Wahrnehmung und Kommunikation*, pp. 29–49. Peter Lang, Frankfurt.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel, Boston, MA.

Maturana, H., & Varela, F. (1987). *The tree of knowledge: The biological roots of human understanding*. Shambhala, Boston, MA.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. W. H. Freeman, Oxford.

McCarthy, J., M.Minsky, Rochester, N., & Shannon, C. (1955). A proposal for the Dartmouth summer research project on Artificial Intelligence. Funding proposal. (First documented use of the term 'Artificial Intelligence'). Retrieved from: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, retrieval date: 20.03.2008.

McClelland, J., Rumelhart, D., & Hinton, G. (1986). The appeal of parallel distributed processing. In Rumelhart, D., McClelland, J., & the PDP Research Group (Eds.), *Parallel Distributed Processing*, Vol. 1, pp. 3–40. MIT Press, Cambridge MA.

Melchner, L. v., Pallas, S., & Sur, M. (2000). Visual behavior induced by retinal projections directed to the auditory pathway. *Nature*, *404*, 871–875.

Merleau-Ponty, M. (2002). *Phenomenology of perception*. Routledge Classics. Routledge, London and New York. Translated by C. Smith. French original published in 1945.

Metzinger, T. (Ed.). (2000). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. MIT Press, Cambridge MA.

Microsoft Corporation (2002). Pointer ballistics for Windows XP. Internet article on product webpage. URL: http://www.microsoft.com/whdc/device/input/pointer-bal.mspx, retrieved 20.03.2008.

Millikan, R. (1984). *Language, thought and other biological categories: New foundations for realism*. MIT Press, Cambridge MA.

Minsky, M., & Papert, S. (1969). *Perceptrons. An Introduction to Computational Geometry*. MIT Press, Cambridge MA.

Morasso, P., Mussa Ivaldi, F., & Ruggiero, C. (1983). How a discontinuous mechanism can produce continuous patterns in trajectory formation and handwriting. *Acta Psychologica*, *54*, 83–98.

Moreno, A., & Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life*, *11*, 161–176.

Nagel, S., Carl, C., Kringe, T., Märtin, R., & König, P. (2005). Beyond sensory substitution: Learning the sixth sense. *J. Neural Eng.*, *2*, R13–R26.

Newell, A., & Simon, H. (1963). GPS: A program that simulates human thought. In Feigenbaum, E., & Feldman, J. (Eds.), *Computers and Thought*, pp. 279–293. R. Oldenbourg KG.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, *370*, 256–257.

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences*. (In press).

Nijhawan, R., & Kirschfeld, K. (2003). Analogous mechanisms compensate for neural delays in the sensory and the motor pathways: Evidence from motor flash-lag. *Current Biology*, *13*, 749–753.

Nijhawan, R. (2004). Motor space, visual space and the flash-lag effect. In Koch, C., Adolphs, R., Bayne, T., Leopold, D., Rees, G., Shimojo, S., Stoerig, P., & Wilken, P. (Eds.), *Proceedings of the 9th annual meeting of the Association for the Scientific Study of Consciousness ASSC9, 24.-27.06.2004, Pasadena, California*. Abstract.

Nolfi, S., & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press, Cambridge MA.

Núñez, R., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*, 401–450.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature*, *398*, 34. Scientific Correspondence.

O'Regan, K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1011.

O'Shea, R. (2004). Psychophysics: Catching the old codger's eye. *Current Biology*, *14*, R478–R479.

Pasemann, F. (1996). Repräsentation ohne Repräsentation: Überlegungen zu einer Neurodynamik modularer kognitiver Systeme. In Rusch, G., Schmidt, S. J., & Breidbach, O. (Eds.), *Interne Repräsentationen - Neue Konzepte der Hirnforschung*, pp. 42–91. Suhrkamp, Frankfurt.

Petitmengin, C. (2005). Un exemple de recherche neuro-phénoménologique : L'anticipation des crises d'épilepsie. *Intellectica*, *40*, 63–89.

Petitmengin, C. (2006). Describing one's subjective experience in the second person. an interview method for the Science of Consciousness. *Phenomenology and the Cognitive Sciences*, *5*, 229–269.

Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (Eds.). (1999). *Naturalizing phenomenology*. Stanford University Press, Stanford CA.

Pfeifer, R., & Scheier, C. (1999). *Understanding Intelligence*. MIT Press, Cambridge MA.

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Delachaux et Niestlé, Neuchátel-Paris.

Piaget, J. (1969). *The child's conception of time*. Routledge & Kegan Paul, London. Translated by A. J. Pomerans. French original published in 1946.

Port, R., & van Gelder, T. (Eds.). (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge MA.

Prinz, J. (2006). Putting the brakes on enactive perception. *Psyche*, *12*, 1–19.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J., & Sosik, P. (Eds.), *Advances in Artificial Life: Sixth European Conference on Artificial Life (ECAL01)*, pp. 357–366. Springer.

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: Long-distance synchronization of human brain activity. *Nature*, *397*, 430–433.

Rohde, M. (2003). Dynamical properties of self-regulating neurons. Bachelor's thesis. Institute for Cognitive Science, University of Osnabrück. BSc in Cognitive Science.

Rohde, M. (2004). Organisation of complex behaviour: A hierarchical modular neural control architecture for the generation of handwriting trajectories. Master's thesis, Department of Informatics, University of Sussex. MSc Evolutionary and Adaptive Systems.

Rohde, M., & Di Paolo, E. (2005). *t* for two: Linear synergy advances the evolution of directional pointing behaviour. In Capcarrere, M., Freitas, A., Bentley, P., Johnson, C., & Timmis, J. (Eds.), *Advances in Artificial Life: 8th European Conference, ECAL 2005, Canterbury, UK, September 5-9, 2005, Proceedings*, Vol. 3630 of *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, pp. 262–271 Heidelberg. Springer.

Rohde, M., & Di Paolo, E. (2006a). Evolutionary Robotics and Perceptual Supplementation: Dialogue between two minimalist approaches. Abstract. 50th Anniversary of Artificial Intelligence Summit, Ascona, Switzerland 9.-14.7.2006.

Rohde, M., & Di Paolo, E. (2006b). An Evolutionary Robotics simulation of human minimal social interaction. Long abstract. SAB'06 Workshop on Behaviour and Mind as a Complex Adaptive System, Rome, Italy 30.9.2006.

Rohde, M., & Di Paolo, E. (2006c). 'Value signals' and adaptation: An exploration in Evolutionary Robotics. Tech. rep. 584, Centre for Research in Cognitive Science, University of Sussex, UK.

Rohde, M., & Di Paolo, E. (2007). Adaptation to sensory delays: An Evolutionary Robotics model of an empirical study. In Almeida e Costa, F., Rocha, L., Costa, E., Harvey, I., & Coutinho, A. (Eds.), *Proceedings of the 9th European Conference on Artificial Life*, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), pp. 193–202 Berlin, Heidelberg. Springer.

Rohde, M., & Di Paolo, E. (2008). Embodiment and perceptual crossing in 2D: A comparative Evolutionary Robotics study. In Tani, J., Asada, M., Hallam, J., & Meyer, J.-A. (Eds.), *Proceedings of the 10th International Conference on the Simulation of Adaptive Behavior SAB'08 in Osaka, Japan 7.-12.7.2008*, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) Berlin, Heidelberg. Springer. (Forthcoming).

Rohde, M., & Gapenne, O. (2006). Proposition de protocole expérimental ARCo'06: Etude de l'adaptation aux délais sensorimoteurs. Submission (winning) to experimental protocol competition. ARCo'06, 6.-8.12.2006 in Bordeaux, France.

Rohde, M., & Stewart, J. (2008). Ascriptional and 'genuine' autonomy. *BioSystems*, *91*(2), 424–433. Special issue on modelling autonomy.

Ross, S. (1984). *Differential Equations* (Third edition). John Wiley & Sons, New York.

Russell, S., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, New Jersey.

Rutkowska, J. (1997). What's value worth? Constraining unsupervised behaviour acquisition. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life EACL97, Brighton UK*, pp. 290–298. MIT Press.

Ryle, G. (1949). *The Concept of Mind*. Penguin Books, London.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417–424.

Seth, A. (2007). Measuring autonomy by multivariate autoregressive modelling. In Almeida e Costa, F., Rocha, L., Costa, E., Harvey, I., & Coutinho, A. (Eds.), *Proceedings of the 9th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence Berlin, Heidelberg. Springer.

Seth, A., & Edelman, G. (2007). Distinguishing causal interactions in neural populations. *Neural Computation*, *19*, 910–933.

Shanon, B. (2001). Altered temporality. *Journal of Consciousness Studies*, *8*, 35–58.

Shemmell, J., Hasan, Z., Gottlieb, G., & Corcos, D. (2007). The effect of movement direction on joint torque covariation. *Experimental Brain Research*, *176*, 150–158.

Smith, K., & Smith, W. (1962). *Perception and motion: An analysis of space-structured behavior*. Saunders.

Smith, R. (2004). Open Dynamics Engine (0.5 release). Retrieved: 10.1.2005. URL: http://ode.org.

Sporns, O., & Edelman, G. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Dev.*, *64*, 960–981.

Steiner, U. (Ed.). (1997). *Husserl*. Philosophie jetzt! Edited by Peter Sloterdijk. Eugen Diederichs Verlag, München.

Stetson, C., Cui, X., Montague, P., & Eagleman, D. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, *51*, 651–659.

Stewart, J. (2004). *La vie : Existe-t-elle?* Vuibert, Paris.

Stewart, J. (2008). Foundational issues in enaction as a paradigm for cognitive science: From the origin of life to consciousness and writing. In Stewart, J., Gapenne, O., & Di Paolo, E. (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*. MIT Press, Cambridge, MA. (In press).

Stewart, J., & Gapenne, O. (2004). Reciprocal modelling of active perception of 2-D forms in a simple tactile-vision substitution system. *Minds and Machines*, *14*, 309–330.

Stewart, J., Gapenne, O., & Di Paolo, E. (Eds.). (2008). *Enaction: Towards a New Paradigm for Cognitive Science*. MIT Press, Cambridge, MA. (In press).

Stilling, N., Weisler, S., Chase, C., Feinstein, M., Garfield, J., & Rissland, E. (1998). *Cognitive Science: An Introduction* (Second edition). MIT Press, Cambridge MA.

Strogatz, S. (1994). *Nonlinear Dynamics and Chaos. With Applications to Physics, Biology, Chemistry and Engineering*. Perseus Books, Cambridge MA.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press, Cambridge, MA.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences*, *4*, 407–427.

Thompson, J., Ottensmeyer, M., & Sheridan, T. (1999). Human factors in telesurgery: Effects of time delay and asynchrony in video and control feedback with local manipulative assistance. *Telemedicine Journal*, *5*, 129–137.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In Bullowa, M. (Ed.), *Before speech*, pp. 39–52. Cambridge University Press, Cambridge.

Tuci, E., Quinn, M., & Harvey, I. (2002). Evolving fixed-weight networks for learning robots. In *Congress on Evolutionary Computation CEC2002 (Proceedings)*, pp. 1970–1975. IEEE Press.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, 615–628.

Varela, F. (1991). Organism: A meshwork of selfless selves. In Tauber, A. (Ed.), *Organism and the origin of the self*, pp. 79–107. Kluwer Academic, Netherlands.

Varela, F. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, *3*, 330–350.

Varela, F. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34*, 72–87.

Varela, F. (1999). The specious present: The neurophenomenology of time consciousness. In Petitot, J., Varela, F., Pachoud, B., & Roy, J.-M. (Eds.), *Naturalizing Phenomenology*. Stanford University Press, Stanford.

Varela, F., Maturana, H., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5*, 187–196.

Varela, F., Thompson, E., & Rosch, E. (Eds.). (1991). *The embodied mind: Cognitive science and human experience*. MIT Press, Cambridge, MA.

Vermersch, P. (1994). *L'entretien d'explicitation en formation initiale et en formation continue*. E.S.F., Paris.

Verschure, P., Wray, J., Sporns, O., Tononi, G., & Edelman, G. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems*, *16*, 247–265.

Vickerstaff, R., & Di Paolo, E. (2005). Evolving neural models of path integration. *Journal of Experimental Biology*, *208*, 3349–3366.

Weber, A. (2003). *Natur als Bedeutung. Versuch einer semiotischen Theorie des Lebendigen*. Königshausen und Neumann, Würzburg.

Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*, 97–125.

Weiss, P., & Jeannerod, M. (1998). Getting a grasp on coordination. *News Physiol. Sci.*, *13*, 70–75.

Welch, R. (1978). *Perceptual Modification: Adapting to Altered Sensory Environments*. Academic Press, New York.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. MIT Press, Cambridge MA.

Wood, R., & Di Paolo, E. (2007). New models for old questions: Evolutionary robotics and the 'A not B' error. In Almeida e Costa, F., Rocha, L., Costa, E., Harvey, I., & Coutinho, A. (Eds.), *Proceedings of the 9th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, pp. 1141–1150 Berlin, Heidelberg. Springer.

Yamauchi, B., & Beer, R. (1994). Sequential behaviour and learning in evolved dynamical neural networks. *Adaptive Behavior*, *2*, 219–246.

Zaal, F., Daigle, K., Gottlieb, G., & Thelen, E. (1999). An unlearned principle for controlling natural movements. *Journal of Neurophysiology*, *82*, 255–259.