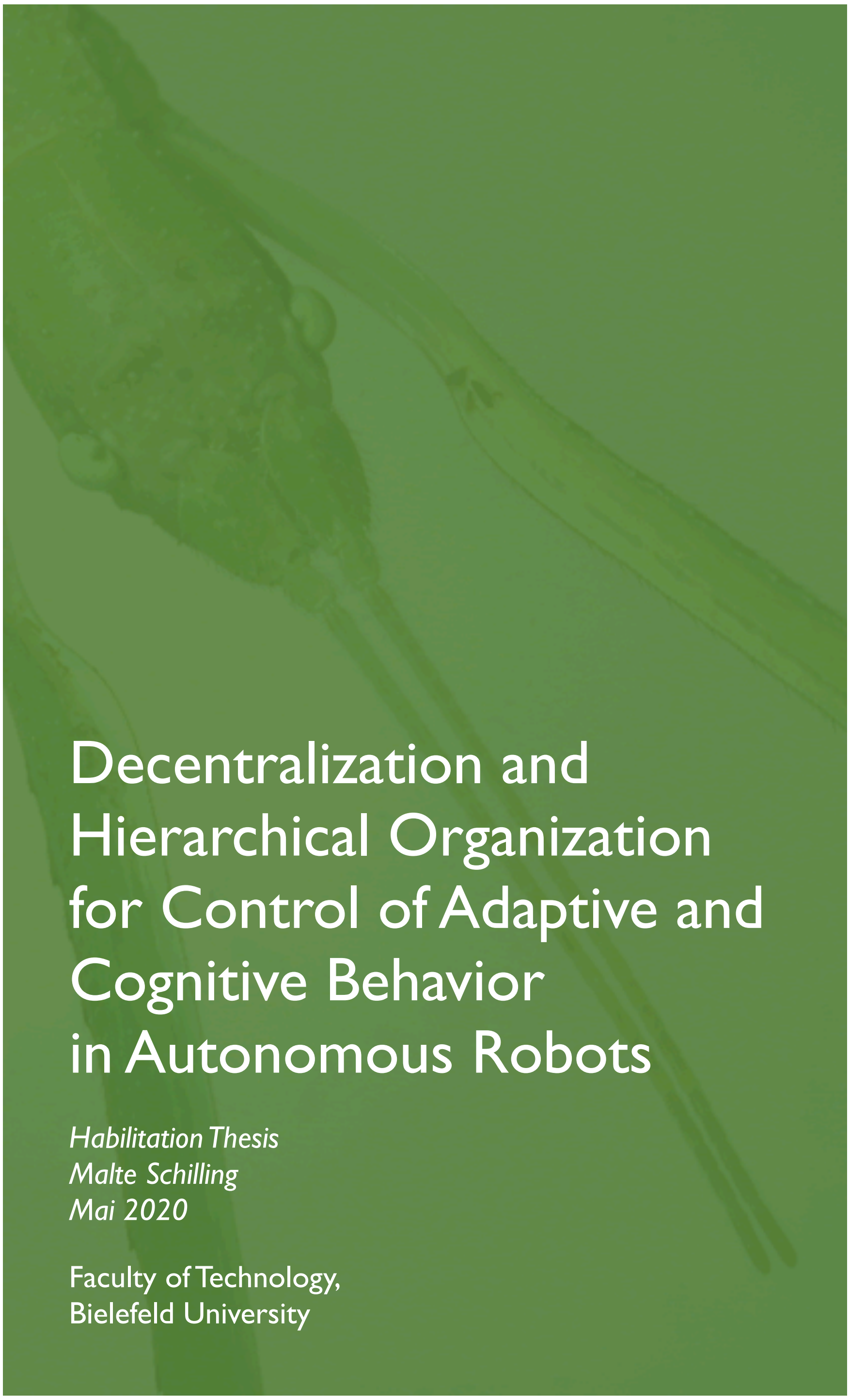


Decentralization and Hierarchical Organization for Control of Adaptive and Cognitive Behavior in Autonomous Robots

Habilitation Thesis

Malte Schilling

Mai 2020

Faculty of Technology,
Bielefeld University

Decentralization and Hierarchical Organization for Control of Adaptive and Cognitive Behavior in Autonomous Robots

Malte Schilling

Bielefeld University, Faculty of Technology

Habilitation Thesis

Bielefeld 2020

Acknowledgements

Even in the field of six-legged walking everything begins with a first step. In my case, what started as a walk has now expanded to a long run. As I am writing up this thesis and bringing together the different parts, I am grateful for everything I learned on my way and all the contributions—even more, all the contributors who walked with me along that way. This thesis is not simply retracing my work, but it literally emerged from all these contributions scattered across disciplines (and continents). Thank you all, for showing me the way!

Many of these stopped being colleagues and became friends along that way. I feel blessed by all the support from these and all my other friends (outside and inside of academia). And even though I rarely reach out and often contact becomes a little scarce—this really means the world to me. Thank you, for tagging along!

There are many people which should be mentioned by name. I can only mention three: Thank you, Holk—even though you are not much of a runner, you put me up for this. I want to thank you for always treating me as a colleague and going along with whatever my direction was. And: I am looking forward for the next step. To my parents: Danke für alles mitgehen, mit auf den Weg bringen und mittragen. Für alles Interesse, aber auch Verständnis für eigene, unverständliche Wege und Themen. It's good to know home, when you are on the run.

Abstract

Cognition—understood as a form of planning ahead—complements adaptive behavior. It leverages knowledge about performing a specific behavior into a novel context while minimizing any harm to the behaving system itself as it is using an internal simulation to predict possible outcomes. In this thesis, I propose a minimal cognitive system that integrates these two kinds of processes in one control system for a six-legged robot. On the one hand, adaptive behavior emerges from interaction of simple local control modules which allows the system to react quickly when facing disturbances. Detailed experimental findings in insects suggests that this evolved flexibility results from a hierarchical and decentralized architecture. While a lower control level coordinates muscle activation patterns and joint movements on a short timescale, a higher level handles action selection on longer timescales. On the other hand, following a bottom-up approach this is extended towards a cognitive system that is able to invent new behaviors and to plan ahead. Using a grounded internal body model planning is realized as a form of internal simulation of possible actions which are applied out of their original context. Exploiting the decentralized architecture this cognitive expansion allows to test and predict properties of newly invented behaviors, while the body is decoupled from the control system.

The thesis introduces the minimal cognitive system as it is applied on the robot Hector in a climbing task. It consecutively introduces the underlying control characteristics and relates these to findings from biology and neuroscience. First, hierarchical organization can be found in many animals and it structures control into parsimonious modules. Second, this is complemented by research on stick insects in particular which offers an even more detailed neuronal and behavioral level for analysis. This emphasizes decentralization of control structures and the importance of an embodied perspective which integrates bodily properties into the concurrent control process exploiting, for example, elasticities of muscles for simplifying the control problem. Third, internal representations are introduced in a bottom-up manner as grounded internal models—realized as recurrent neural networks—that are at first considered in the context of serving a specific behavior. Fourth, as a consequence, cognitive processing is realized as recruitment of the already existing flexible internal models in an internal simulation. The underlying architecture is applied on the hexapod robot Hector and analyzed in detail in simulation. Furthermore, learning is considered for this approach.

Contents

Abstract **v**

Preface **xi**

1 Introduction **1**

1.1 Adaptive Behavior 2
1.2 Cognitive Behavior 4
1.3 Dual Process Perspective 5
1.4 Overview of Control Architecture 7

2 Decentralization and Modularization – Motor Control Hierarchies **13**

2.1 Embodiment 14
2.2 Hierarchical Organization in Biological Motor Control 15
2.3 Hierarchical Organization in Technical Approaches 17
2.4 Decentralization in Biological Motor Control 18
2.5 Walknet – a Decentralized Control Architecture 19
2.6 Results Overview 21
2.7 Conclusions 24
2.8 List of Publications 25
2.8.1 Contributions to the Thesis 25
2.8.2 Further Related Publications 26

3 Learning of Adaptive Behavior in a Decentralized Fashion **29**

3.1 Hierarchical Representation in Robotics 30
3.2 The Advantage of Decentralized Control 32
3.3 Decentralized Deep Reinforcement Learning 36
3.3.1 Reinforcement Learning 36
3.3.2 Decentralized Motor Control Architecture 37
3.3.3 A Centralized Approach as a Baseline 38
3.3.4 Policy Network 38
3.4 Results Overview 38
3.5 Conclusions 40

3.6 List of Publications . . . 42
3.6.1 Contributions to the Thesis . . . 42
3.6.2 Further Related Publications . . . 43

4 Hierarchical Internal Body Models — Grounding in Adaptive Behavior 45
4.1 From Adaptive to Cognitive Behavior . . . 45
4.2 Internal Body Models . . . 46
4.3 Mean of Multiple Computation Principle . . . 48
4.4 Extension towards a Hierarchical Body Model . . . 48
4.4.1 Application for Hexapod Locomotion . . . 50
4.4.2 Application in Bimanual Reaching Task . . . 52
4.5 Conclusions . . . 52
4.6 List of Publications . . . 53
4.6.1 Contributions to the Thesis . . . 53
4.6.2 Further Related Publications . . . 54

5 Embodied Internal Simulation as Planning Ahead 57
5.1 Recruitment of Grounded Internal Models . . . 57
5.2 Structure of the Cognitive Expansion . . . 58
5.3 Results Overview . . . 63
5.4 Conclusions . . . 67
5.5 List of Publications . . . 68
5.5.1 Contributions to the Thesis . . . 68
5.5.2 Further Related Publications . . . 69

6 Discussion and Conclusion 71
6.1 Key Principles of Hierarchical Control . . . 72
6.1.1 Lower Level Motor Control . . . 73
6.1.2 Mid-level Action Control . . . 75
6.1.3 Higher Level Control – Planning . . . 76
6.2 Conclusion . . . 77
6.3 List of Publications . . . 80
6.3.1 Contributions to the Thesis . . . 80
6.3.2 Further Related Publications . . . 81

Appendices

Overview Articles in Appendix 87

A Biological-inspired Locomotion Control Articles **93**

A.1 Walknet, a bio-inspired controller for hexapod walking 94

A.2 Decentralized control of insect walking – a simple neural network explains a wide range of behavioral and neurophysiological results . 118

A.3 A hexapod walker using a heterarchical architecture for action selection 168

B Decentralized Learning Articles **187**

B.1 From Crystallized Adaptivity to Fluid Adaptivity in Deep Reinforcement Learning 188

B.2 Decentralized Deep Reinforcement Learning for a Distributed and Adaptive Locomotion Controller of a Hexapod Robot 196

B.3 An Approach to Hierarchical Deep Reinforcement Learning for a Decentralized Walking Control Architecture 206

C Hierarchical Internal Model Articles **219**

C.1 Grounding an Internal Body Model of a Hexapod Walker – Control of Curve Walking in a Biological Inspired Robot 220

C.2 What's next: Recruitment of a grounded predictive body model for planning a robot's actions 228

C.3 Hierarchical Dual Quaternion-Based Recurrent Neural Network as a Flexible Internal Body Model 248

D Articles on Planning Ahead in a Cognitive Architecture **259**

D.1 From Adaptive Locomotion to Predictive Action – Cognitive Control for a Six-Legged Walker 260

D.2 ReaCog, a Minimal Cognitive Controller Based on Recruitment of Reactive Systems 276

D.3 Old Actions in Novel Contexts – a Cognitive Architecture for Safe Explorative Action Selection 300

E Further Perspective Articles **309**

E.1 Avoid the hard problem: Employment of mental simulation for prediction is already a crucial step 310

E.2 Learning by seeing – associative learning of visual features through mental simulation of observed action 312

E.3 Simulation across Representation: The Interplay of Schemas and Simulation-Based Inference on Different Levels of Abstraction . . . 322

Bibliography **327**

x

Preface

This habilitation thesis presents the author's selected findings on the organization, realization, and learning of motor control structures and cognitive function applied on a six-legged walking robot. It is written as a cumulative thesis which consists of two parts. The main scientific contributions have been published in 15 papers. Two articles have been recently submitted, all other articles were accepted following peer-review. Seven have been published as a journal publication (mean impact factor of 5.65) and six have been published at international conferences (either high ranked CORE A conferences or specialized conferences on specific topics). For all 15 articles, I am the first author (further detailed declaration of author contributions are given in the specific chapters and have been acknowledged by the other authors). These articles constitute the second and major part of the thesis (further related publications are pointed out in the chapters as well).

The first part of the thesis introduces the main findings. It is meant as a self-contained text which, on the one hand, summarizes the articles and, on the other hand, puts these into the broader context of my research program pointing out how each article contributes to our understanding of adaptive behavior and cognitive function in autonomous systems. As a consequence, this first part draws on the content of the published articles: figures, results, and arguments are selected from the original publications and explained with respect to the overarching line of thought. Further used sources are mentioned accordingly.

I used to think that the brain was the most wonderful organ in my body.
Then I realized who was telling me this.

— Emo Philips

We believe that the basic function of cognition is control of action. From an evolutionary perspective, it is hard to imagine any other story.

— Glenberg & Gallese (2012, p. 918)

1 Introduction

Contents

1.1 Adaptive Behavior **2**
1.2 Cognitive Behavior **4**
1.3 Dual Process Perspective **5**
1.4 Overview of Control Architecture **7**

In animals, adaptivity characterizes a form of behavior that shows a robustness even when facing varying environmental conditions. Adaptive behavior allows a system to deal flexibly with the unpredictability of the environment, but it is tied to a specific context. Cognition—understood as a form of planning ahead—complements adaptive behavior (McFarland & Bösser, 1993) as a form of dealing with novel contexts. It is leveraging existing knowledge on performing a behavior into a novel context, minimizing any harm to the system itself as it is using an internal prediction of possible outcomes.

These two kinds of behaviors—that are widely spread in humans and animals—are assumed to rely on two distinct types of processes subserving behavior. On the one hand, automatic and effortless processes allow to quickly adapt to changes in the environment and, on the other hand, a reflective and controlled process deals with planning ahead as a form of internal simulation. One important question concerns how these processes are associated. In this thesis, we address the nature of these two types of processes from the perspective of Embodied Cognition (Barsalou, 2008) which assumes that these processes are highly intertwined. The thesis will analyze control characteristics of adaptive behavior and cognitive behavior as well as how these are interlinked. Overall, the goal is to understand these characteristics and defining control principles. This is approached through a constructive modeling approach that starts from detailed findings in neuroscience, biology, and cognitive science and aims at realizing a functioning minimal cognitive system. Following such a bottom-up approach the detailed goals of this thesis are:

- Delineate key characteristics of adaptive behavior and the underlying control system from experiments and analysis on walking of stick insects.
- Realize hierarchical organization and decentralized organization as such key mechanisms in an artificial control system that can be simulated and run on a real robot which showcases the effectiveness of these mechanisms.
- Establish decentralization as an advantageous principle for emergent adaptive behavior in a comparative (deep reinforcement) learning study.
- Show how internal models—as a key component of cognition—are grounded in adaptive behavior making them a necessary condition for certain types of coordinated adaptive behavior.
- Describe cognitive behavior as a form of internal simulation and provide an overview of the neuroscientific findings supporting such an Embodied Cognition view.
- Extend the existing adaptive control structure towards a cognitive system that allows to recruit the grounded internal models in a form of mental simulation. This will provide a proof of concept for how an adaptive system can leverage its' existing control structure towards novel situations and establish a transfer towards novel contexts. Such a system constitutes a minimal cognitive system.

The introduction will briefly introduce characteristics of adaptive and cognitive behavior as well as relate these towards dual process theories. The goal is to give the reader a broad overview of topics covered. This thesis will mostly deal with locomotion as an example which will be motivated before an outline of the developed control architecture will be given.

1.1 Adaptive Behavior

Adaptive behavior in animals deals with reacting to broad variations imposed by the particular environmental niche. Consider, for example, an insect that is able to climb through a twig even though there is only very limited information on possible footholds and it is impossible to predict the movements of the substrate. Following Beer and colleagues (Beer, 1990; Beer *et al.*, 1990) adaptive behavior can be broadly defined as

> *"behavior [that] is continuously adjusted to meet the ever changing internal and external conditions of the interaction."* (Beer *et al.*, 1990, p. 171) or from a more functional perspective: *"in other words, adaptive behavior is the result of the continuous interaction between the nervous system, the body and the environment, each of which have rich, complicated, highly structured dynamics."* (Chiel & Beer, 1997, p. 555)

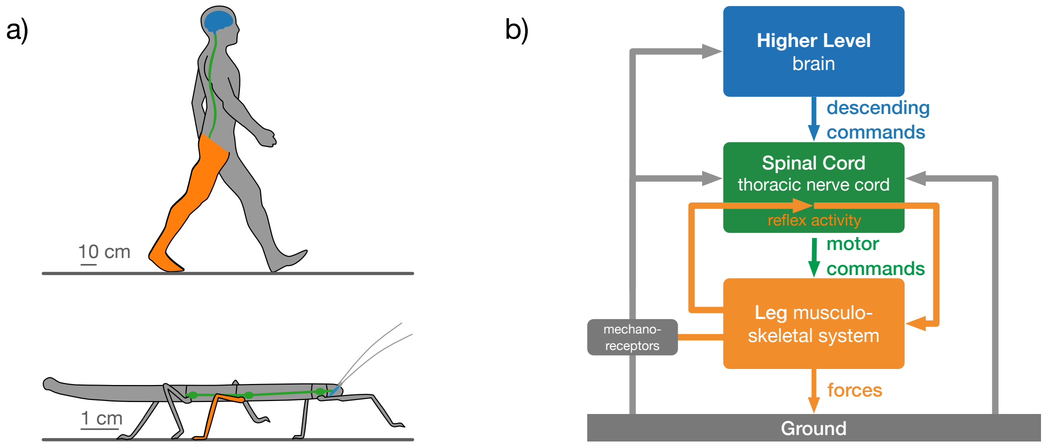


Figure 1.1: Hierarchical organization in motor control. A general hierarchical structure of motor control is shared in humans and animals (Dickinson *et al.*, 2000). This is shown in a); colors signify different levels of this motor hierarchy: higher level is shown in blue, an intermediate (and decentralized) control level is shown in green. Interactions with the environment (including preflexes and properties of muscles) are shown in orange. Color coding applies to the whole thesis. b) shows a simple schematic of this organization. Research on humans mostly focus on the higher levels shown in blue (Magill & Anderson, 2017) and is rarely extended towards the intermediate level (Arber & Costa, 2018). This is nicely complemented by work on animals and insects, in particular, that adds complementary findings on an embodied level (shown in orange) that highlights interaction with the environment (Dickinson *et al.*, 2000). The simplified schematic combines these approaches which is a starting point for the considerations of the presented architecture (for a complete schematic see Fig. 1.4).

Such traits of adaptivity are tried to be captured and exploited in technical systems. But even today most technical systems appear brittle in this regard as they follow an automation paradigm (Lipson, 2019; Hauser, 2019). Technical systems still have difficulties facing noisy environmental settings and cannot handle even slight changes in the appearance or configuration of an environment. This includes many of the current reinforcement learning approaches that aim to solve a specific problem in a static context (for review see (Neftci & Averbeck, 2019); or see (Finn *et al.*, 2017) highlighting problems when quick adaptations are required). In contrast, biologically-inspired control approaches try to address these problems and mimic how animals deal with such disturbances. These approaches aim to uncover the underlying structure of the variability or to characterize it in probabilistic terms. Such forms of adaptivity allow for optimization of a controller with respect to a given natural environment as a typical test and target domain, which has led to stable control approaches for different behaviors (Billard & Kragic, 2019; Cully *et al.*, 2015; Fazeli *et al.*, 2019; Hwangbo *et al.*, 2019). While this form of adaptivity leads to approaches that are well tuned to specific (narrow) contexts and the characteristics of these environments, such approaches are often prone to overfitting, too. It still does not allow to adapt towards broader variations of the environment.

One approach to deal with changing environmental conditions is to explicitly distinguish between different contexts and use specific control structures for each different situation. This induces a form of hierarchical organization (Binder *et al.*, 2009; Botvinick, 2008) in which behaviors that are realized as lower level control primitives are selected at a higher level (Flash & Hochner, 2005; Schaal & Schweighofer, 2005). Such control hierarchies have been inspired by the organization of motor control systems in humans and animals (Fig. 1.1). Importantly, there is a considerable difference between adaptivity at the different levels of the control hierarchy. The lower level allows for gradual fine-tuning and optimization because the relationship between changes in the control parameters and defined performance metrics can be assumed to stay more or less smooth in a given context. This is ideal for (gradient-based) learning approaches. In contrast, the higher level deals explicitly with switching between different behaviors, and we cannot expect smooth transitions between these behaviors. This complicates learning and realizing of such hierarchical control approaches. Furthermore, the different levels operate on different timescales. While on the lower level motor control tasks require fast reactions and adaptations, higher level selection of behaviors depending on the current context should be more stable (Namikawa *et al.*, 2011).

As selection of behaviors explicitly aims at executing a behavior in a specific context, it implicitly induces a conceptualization of the space of possible contexts and applicable behaviors (Hay *et al.*, 2018). Adaptivity on this longer timescale allows for dealing with different forms of environments and selecting an appropriate response. Such a control approach is based on the notion that environments can differ in a way that requires understanding them differently and that requires differentially behaving control systems. Assuming that it is beneficial to distinguish environments that appear distinctively different to a system, this introduces the possible problem that an animal or system is encountering a novel environmental situation for which there is no appropriate and optimized adaptive behavior as it was never experienced before. One solution is to always resort to one behavior (that could be considered a default behavior). This appears to be a solution we find in many biological systems and which has been transferred to robots as well (Shamsuddin *et al.*, 2011).

1.2 Cognitive Behavior

Cognition offers another, additional solution. While action selection tries to narrow down which selection appears most applicable (from experience or as an evolved trait), cognition carries the notion of becoming 'creative' (Bongard & Lipson, 2014). This means, that during cognitive selection of an action the scope of possibly applicable behaviors is widened and behaviors are tested outside their original, defining niche. Such a trial-and-error approach entails risks as a behavior is applied in a novel context and the consequences have not been experienced before. But the consequences of applying the behavior are not necessarily random. In the case of body movements, for example, they are predictable. Evaluating predictions of consequences of behaviors beforehand is what constitutes an internal simulation—it provides a 'what-if' mechanism (Lake *et al.*, 2017).

In this context, cognitive behavior is meant in the following way:

> Cognition can be understood as the ability to plan ahead (McFarland & Bösser, 1993) by means of an internal simulation (Hesslow, 2002) relying on an internal representation (Glenberg, 1997; Lake *et al.*, 2017; McNamee & Wolpert, 2019)—starting with a model describing the spatial and dynamic relations (Acosta-Calderon & Hu, 2005) of the own body (Cruse, 1999)—which is grounded in embodied experiences (Gallese & Lakoff, 2005; Steels, 2003).

From such a perspective, cognition as a mechanism aims in a different direction than fitting better and better to a certain niche. Internal simulation as a principle allows to react when there is no more stationary niche, but it becomes necessary to constantly reevaluate potential behavioral alternatives when facing novel contexts (Broekens, 2005; Hassabis *et al.*, 2017).

1.3 Dual Process Perspective

There is a long tradition for such a distinction into two qualitatively different kinds of systems or processes with complementary advantages (Schneider & Chein, 2003) which can be traced back to Schneider & Shiffrin (1977). Many dual process or dual system theories have been proposed (for a review see (Stanovich & West, 2000; Evans & Stanovich, 2013)) which usually distinguish between automatic and controlled processing. Most prominently this has been formulated as a two systems theory by Kahneman (2011). Here, we will use the notion of types of processes following Evans (2008) who pointed out that there can be a multitude of such processes and that many of these theories address or highlight different characteristics and not deal with distinct systems (this notion is in general in agreement with Kahneman & Frederick (2002) who understand a system as a collection of processes). Evans (2008) distinguishes type 1 processes that are fast and automatic in contrast to type 2 processes that are slow and effortful. The two different types of processes are linked to explain more habit-based behavior (well learned acts (Norman & Shallice, 1986)) respectively cognitive behavior (Schneider *et al.*, 2020).

Automatic type 1 processes have been originally defined as activation of behavior that *"becomes active in response to a particular input configuration"* and that *"is activated automatically without the necessity for active control or attention by the subject"* (Schneider & Shiffrin, 1977, p. 2). Such processes are assumed to be autonomous (Dickinson, 1985; Evans & Stanovich, 2013) and allow to act as well as react on a fast timescale (for further characteristics see Fig. 1.2 which details characteristics of the two types of processes). Importantly, a fast response is widely assumed to rely on parallel processing (Evans, 2008). Therefore, this is assumed as one defining characteristic for type 1 processes which fits well to what we find in biology. Parallel processing is not only found in the brain, but extends over the whole nervous system in which, furthermore, processing often occurs locally, for example, in reflex-like behavior.

In contrast, type 2 processes are defined by the ability of flexibly decoupling internal models outside of their original context (Evans & Stanovich, 2013). Together

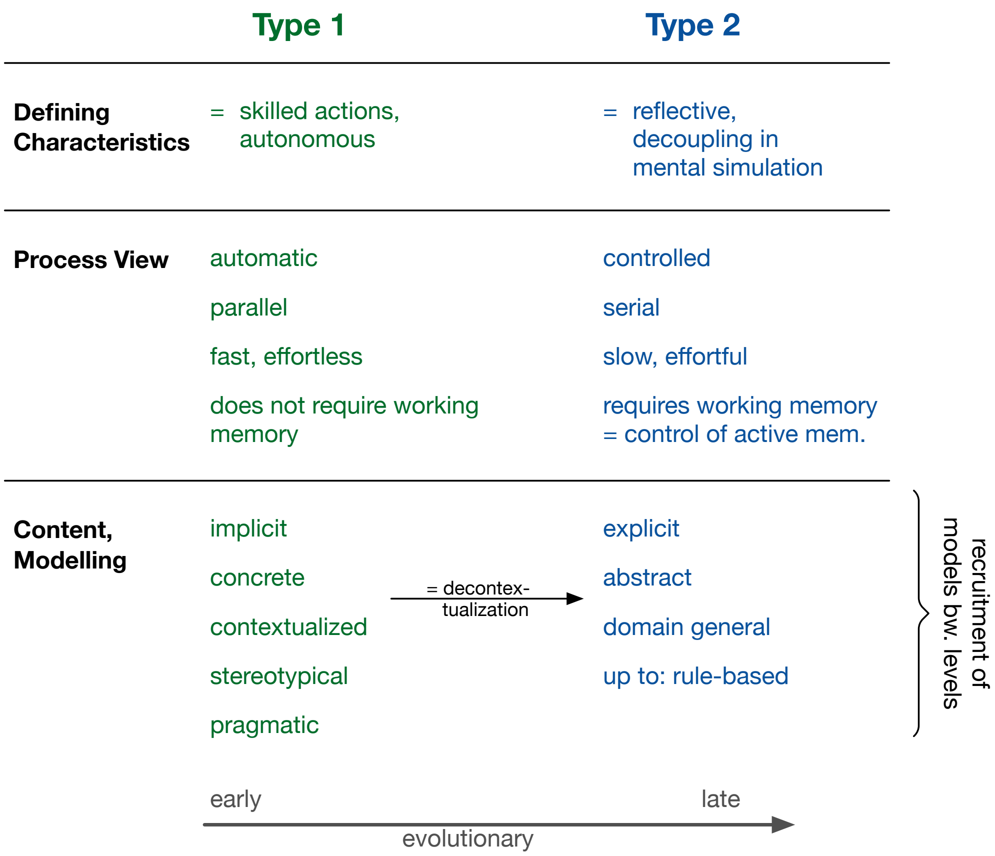


Figure 1.2: Dual process characteristics: Overview of features associated with the two different types considered in dual process or dual system theories. Higher level, type 2 process characteristics are shown again in blue. Type 1 process characteristics are shown in green. Defining characteristics and process features follow largely Evans & Stanovich (2013) which provides an overview of dual process accounts that originated from (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). This is extended by a perspective on contents on which these processes operate on and their relation to internal modeling. The nature of these internal models is based on Kahneman & Frederick (2002).

with predictive capabilities this flexible use (Schneider *et al.*, 2020) of internal models constitutes an internal simulation (and for Evans & Stanovich (2013) represents a key feature of working memory). More broadly speaking and going back to the original definition, type 2 processes are related to behavior *"under control of, and through attention"* which is *"tightly capacity limited, but the costs of this capacity limitation are balanced by the benefits deriving from the ease with which such processes may be set up, altered, and applied in novel situations for which automatic sequences have never been learned"* (Schneider & Shiffrin, 1977, p. 2, 3). For more details on the distinction see Fig. 1.2.

Type 2 processes are used in cognitive behavior dealing with a novel situation. But while the two different types of processes address different types of tasks, it is important to note that these are tightly interconnected (Schneider *et al.*, 2020). Cognitive behavior involves recruitment of underlying type 1 processes in a mental simulation (Anderson, 2010). This notion of recruitment is also present in dual process approaches. For example, in Norman & Shallice (1986) a type 2 cognitive control system acts as a 'supervisory attentional system' that activates and recruits the underlying type 1 habit-based system during cognitive behavior. A further connection between the systems is given through learning: cognitive behavior that has been used in a novel context can be entrained as a skill into system 1 (Kahneman & Frederick, 2002).

Often, type 2 processes—or a system 2—have been related to higher level competencies as are language or consciousness and which are unique to humans. But importantly, there is broad support for many of the characteristics of both types of processes in many animals—including insects (Menzel *et al.*, 2007; Giurfa & Menzel, 2013)—that includes as well a distinction into these two different types of processes (Evans, 2008).

Such a distinction into two types of processes is advantageous (Schneider & Chein, 2003) as it allows, on the one hand, for fast, robust, and adaptive behavior. On the other hand, a higher level for controlled processes extends this to cognitive behavior that allows to safely plan and apply a behavior outside of its original context and transfer skills between situations. This is a prerequisite for a form of social learning through observation of others or through instruction using language which both rely on a controlled reactivation of underlying processes and representation. A dual process view is also well supported by neuroscientific findings on two types of system involved in behavioral decisions (Daw *et al.*, 2005) which are tightly connected to reinforcement learning and the distinction of model-free and model-based learning (Niv, 2019; Neftci & Averbeck, 2019; Botvinick *et al.*, 2019; Lee *et al.*, 2019).

1.4 Overview of Control Architecture

Both types of processes will be integrated in the derived architecture which follows a constructive bottom-up approach. The goal is to apply biologically-inspired principles in a functioning control architecture for a robot. The application in a real technical system in simulation and on a real robot demonstrates the reach and effectiveness of these principles as well as it provides opportunities for detailed analysis on different levels, as on a behavioral or neurophysiological level. Specifically, we focus on six-legged walking as an example. Locomotion provides a prime example for adaptive behavior as there are rich interactions with the environment that require fast reactions and coordinated movements of a large number of actuators (Dickinson *et al.*, 2000). Animals excel at walking and running behavior as they can adjust to complex terrains, cluttered environments and all kinds of disturbances, including injuries. These properties make animals interesting models for control approaches as they show a working system that can handle variability and produce stable as well as adaptive behavior (Hwangbo *et al.*, 2019).

We turn towards insects as these provide good model systems to study general motor control principles in detail (Webb, 2020). Insects have accessible and tractable

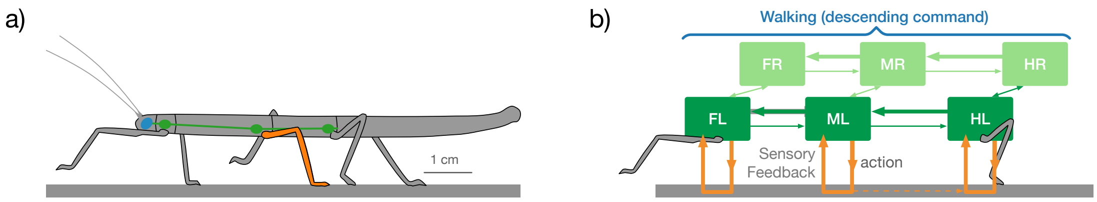


Figure 1.3: Decentralized motor control structure: a) neuroanatomic organization in the stick insect. b) schematic of the decentralized organization of the stick insect control system which is used as a model for the six-legged robot. Colors signify different levels of the motor hierarchy: higher level is shown in blue (as, for example, information on walking direction). Local leg control level is shown in green (Front, Middle, Hind leg on Right and Left side) with coordination influences between neighboring legs shown as arrows. Interaction with the environment (including preflexes and properties of muscles) are shown in orange. Behavior emerges as a result of decentralized and locally interacting concurrent control structures (Schilling *et al.*, 2013*a*).

nervous systems, yet they produce complex motor behaviors that they adjust to changing environmental conditions (Krakauer *et al.*, 2017; Ritzmann & Büschges, 2007; Tuthill & Wilson, 2016). Walking with six legs itself already poses a quite hard problem. Typically, an insect leg consists of three main joints producing the movement of a leg. Each leg pair is connected to one body segment, and even if the possible movements between body segments are restricted, this leads to 18 degrees of freedom overall that have to be controlled by the system. Therefore, the system is highly redundant as the movement of a single joint is influenced by the movements of all joints of the other (standing) legs (Bernstein, 1967). Coordination of the single joint movements of the standing legs through one monolithic system appears problematic and only for specific joint configurations tractable at all.

The goal for the control system (shown in Fig. 1.4) is to demonstrate adaptive behavior—as in walking through uneven terrain—and to allow for cognitive behavior when facing novel problems, for example, crossing a gap that requires adjusting foot positions of the robot. We follow a bottom-up approach starting from the lower level and working upwards towards the higher levels. The following chapters will contribute to the overall architecture:

2. **Decentralization and Modularization – Motor Control Hierarchies:** The next chapter will introduce a decentralized control system for a hexapod walking system. It will summarize detailed findings from biology on walking in insects and derive principles that are implemented in the Walknet system (see Fig. 1.3). The focus is on two key characteristics that can be found throughout the animal kingdom: a hierarchical organization of motor control systems and decentralization. A decentralized organization allows for fast reactions and local behavioral decisions. Such a structure is inherently processing in parallel and allows for fast actions and reactions. The overall behavior emerges out of the interaction of the local processing modules. In this way, the presented

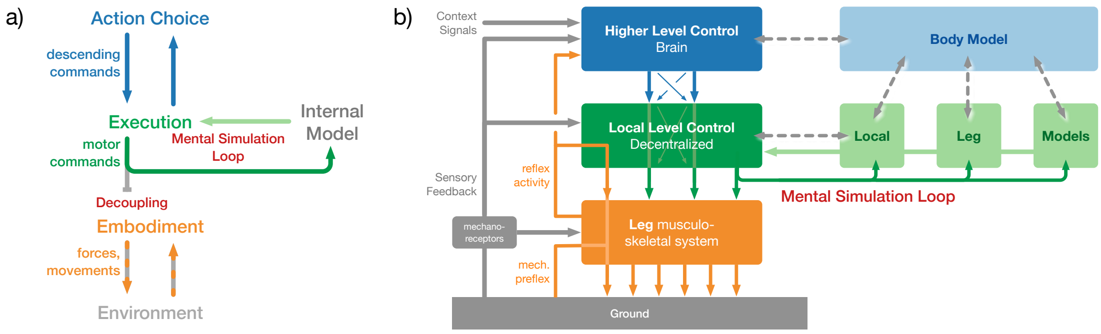


Figure 1.4: Overview of the motor control architecture: The architecture realizes adaptive and cognitive behavior. a) provides an abstract and simplified schematic that visualizes processes between different levels of hierarchies in adaptive behavior (for more details see the following chapter). This realizes type 1 processes (shown on the left). On the right part of a), the connection towards internal models is visualized. This highlights how in a type 2 process internal models are recruited and utilized for internal simulation while the real body is decoupled from the control system. b) shows an overview of the motor control hierarchy and how decoupling of the body allows to realize a form of planning ahead as mental simulation. This constitutes cognitive behavior as a type 2 process. Importantly, there are highly parallel connections which highlight the concurrent and decentralized structure of the system. On the right, in light shaded colors, underlying internal models are shown that serve behavior. Gray dashed arrows between the two columns signify that internal models are grounded and recruited in motor control. The single green arrow to the right indicates that during mental simulation the motor control information is rerouted towards these predictive internal models starting a mental simulation. This schematic will be derived in detail throughout this thesis.

system realizes type 1 processes on this level (Fig. 1.4 a) shows an abstract processing schematic).

The system consists of a hierarchical organization (see Figures 1.1 1.3, 1.4: higher levels will be—throughout this thesis—color coded in blue, an intermediate and more decentralized level in green) which is in agreement with neuroscientific findings on the organization of the motor system in animals and humans (Dickinson *et al.*, 2000; Botvinick, 2007).

Following the paradigm of Embodied Cognition, we are interested in a real system interacting through its body with the environment: it is embodied. As an advantage this allows to exploit mechanical properties of the body as, for example, elasticities of muscles. While the notion of embodiment is not considered in theoretical accounts of dual systems, it has to be integrated in a real control architecture and is now considered as an integral challenge for robotics (Yang *et al.*, 2018). This will be further laid out in the second chapter.

Overall, the second chapter introduces the Walknet architecture and demonstrates the adaptivity of the system in different locomotion tasks.

3. **Learning of Adaptive Behavior in a Decentralized Fashion:** The third chapter will analyze in particular the influence of a decentralized organization on learning. While hierarchical organization is now employed in first Deep Reinforcement Learning approaches in simple, game-like scenarios, this chapter will show how a local and decentralized structure of the motor control architecture facilitates learning of adaptive behavior: it leads to faster convergence and better performance when applied for a simulated hexapod robot. This appears as a promising direction in general for application of Deep Reinforcement Learning in motor control tasks and leverages biological principles into such learning approaches (Hassabis *et al.*, 2017).

4. **Hierarchical Internal Body Models:** Central to cognitive behavior is the notion of internal models. As mentioned above, we assume that in cognitive behavior higher level control processes (type 2) recruit underlying processes and internal models which are grounded in lower levels (Barsalou, 2008). The fourth chapter introduces a flexible and hierarchical internal body model as a recurrent neural network. While the size of this recurrent neural network model is still quite small, it can be employed in motor control serving coordination of leg movements during walking. In addition, it is predictive and allows to be utilized in internal simulation for planning ahead. The chapter briefly summarizes the function of internal models in motor control in general and introduces the recurrent neural network approach for such a body model which is extended towards a hierarchical model.

5. **Internal Simulation as Planning Ahead:** The fifth chapter will extend the Walknet architecture towards a cognitive architecture that allows for planning ahead realized as a form of mental simulation. In mental simulation, the body of the system will be decoupled from the (intermediate and higher level) controller and instead a behavior will be tested out of its original context using predictions of the internal body model. These predictions allow the system to determine if a behavior provides a non-dangerous and suitable solution to the problem at hand. The higher level processing realizes a type 2 process and in particular fulfills the defining characteristics discussed above (Fig. 1.4 a) shows an abstract processing schematic highlighting the decoupling and mental simulation loop during planning ahead). The cognitive expansion will be introduced in detail and the whole system will be, on the one hand, applied on the real robot Hector and, on the other hand, systematically tested in dynamic simulations showing the adaptivity and flexibility of this approach for a task as climbing through an environment with uncertain footholds.

6. **Discussion and Conclusion:** The thesis will conclude with a summary of key characteristics of adaptive and cognitive control systems and provide an outlook on how such a minimal cognitive system might be further extended towards a cooperative and learning system that could handle complex sequential tasks even when requiring collaboration.

The whole architecture represents a minimal cognitive system. Overall, behavior emerges from the interplay between the physical realization of the system embedded in the environment, decentralized local mechanisms—especially on the lower level—and higher level mechanisms in selection of behavior or planning of cognitive behavior. One particular focus is on the complementary role of internal models that span both type of processes as they are grounded in lower level adaptive behavior and recruited for planning ahead in mental simulation by higher level processes.

Biorobots are becoming important scientific tools and can be used to investigate locomotion and to test hypotheses about the underlying interactions of body, control, and environment.

— Ijspeert (2014, p. 196)

2 Decentralization and Modularization – Motor Control Hierarchies

Contents

2.1 Embodiment **14**
2.2 Hierarchical Organization in Biological Motor Control **15**
2.3 Hierarchical Organization in Technical Approaches . . **17**
2.4 Decentralization in Biological Motor Control **18**
2.5 Walknet – a Decentralized Control Architecture **19**
2.6 Results Overview **21**
2.7 Conclusions **24**
2.8 List of Publications **25**
2.8.1 Contributions to the Thesis 25
2.8.2 Further Related Publications 26

Adaptive behavior allows animals to produce stable behavior in broad contexts. They flexibly adapt to changing specific environmental conditions and can deal with variations under these conditions even when facing uncertainty. This robustness to uncertainty still sets animal behavior apart from current engineering and learning solutions. In this chapter, we will motivate our architecture for hexapod walking and summarize the main contributions of our work described in the accompanying publications. The goal is to highlight control principles found in biology and transfer these into a technical system that can be used to control a six-legged robot. Following a bottom-up approach, first, the notion of embodiment is introduced and shown how exploiting physical properties can simplify motor control. Next, we will take inspiration from the organization of motor control in animals with a focus on two key insights: On the one hand, hierarchical organization—actions can be decomposed into sub-actions on different levels of a hierarchy which allows for flexible recombination and induces temporal abstraction (Binder *et al.*, 2009; Mengistu *et al.*, 2016; Uithol *et al.*, 2012; Haruno *et al.*, 2003). On the other hand, and as a key characteristic setting this approach apart, we focus on the modular and decentralized structure of the motor control system. Modularity of a network is understood as an organization in which a network consists of *"multiple densely connected clusters, each with only a limited connection to other clusters"* (Ellefsen *et al.*, 2020, p. 3). With respect to biological and brain-inspired approaches, we specifically use the term decentralization which explicitly points out that this control structure encompasses the whole nervous system and control is distributed to local control modules. Decentralization describes the general idea of concurrent modules that allow for fast, local computations realizing, for example, reflex-pathways (Clune *et al.*, 2013) as found in animals and humans (Alon, 2006; Mountcastle, 1997). Overall behavior emerges from the interaction of the parallel and distributed processing of information along the motor control hierarchy, tying these two characteristics tightly together. These principles will be applied in our Walknet system. Finally, we will demonstrate the adaptivity of this system.

2.1 Embodiment Exploiting Properties of the Body

Adaptive behavior not only depends on neural circuits, but also on the interaction of the body and the environment (Chiel & Beer, 1997; Chiel *et al.*, 2009; Nishikawa *et al.*, 2007). This has been well demonstrated in insects (see Figure 2.1 a) and b)): Legged locomotion in insects spans a behavioral continuum from slow walking to fast running. On the one hand, during slow walking—which will be the main focus of this chapter—insects place their legs accurately in space based on detailed sensory information about the body and the environment (Niven *et al.*, 2012; Theunissen *et al.*, 2014). This is critical when traversing cluttered environments such as canopies, in which secure footholds are sparse. During fast running, on the other hand, the effectiveness of sensory feedback might be constrained by sensorimotor delays (More & Donelan, 2018). For example, at top running speed, a cockroach lifts each of its six legs 20 times per second, corresponding to a step period of only 50 ms (Full & Tu, 1991). This might not leave enough time for sensory feedback to adjust leg movements on a step-by-step basis (Jindrich & Full, 2002; Zill & Moran, 1981). Therefore, it is assumed that fast running in insects is driven predominantly by central oscillating units (Figure 2.1 b) shown in green) in a feed-forward fashion (Bidaye *et al.*, 2018). A shift towards more feedforward control in (fast) locomotion is assumed in other animals as well (Clancy *et al.*, 2019) (Fig. 2.1 c)) and such feedforward control has been applied to multiple legged robots (Ijspeert, 2008). Importantly, insects can still recover from perturbations such as uneven terrain during fast running (Jindrich & Full, 2002; Sponberg & Full, 2008). This is enabled by passive forces from the musculo-skelatal system (Ache & Matheson, 2013; Dudek & Full, 2006), which act more quickly than sensory reflexes as mechanical "preflexes" (Brown & Loeb, 2000) (shown in orange in Figure 2.1 a) and b).

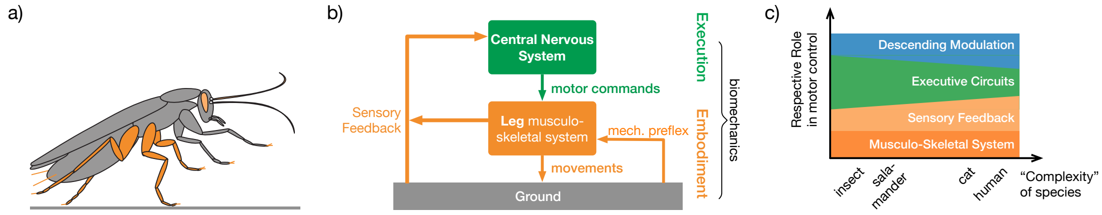


Figure 2.1: Contributions to hierarchical motor control: On the one hand, shown for insects (Grillner, 2003) as such simpler model systems allow for more detailed analysis of interaction between body and the environment. Experiments in insects have stressed the importance of very fast and local reflex activity which is controlled directly on the lowest level or even realized by passive properties as are muscle elasticities or preflexes (shown in orange). Figure a) and b) are adapted from (Dickinson *et al.*, 2000), with a) showing a fast running cockroach highlighting sensors and actuators that are in interaction with the environment (in orange). In b) a general scheme is given, highlighting that neural and mechanical feedback play roles in the control of locomotion: on the shown intermediate level (green), the central nervous system produces motor commands. These activate the musculo-skeletal system of the animal which acts on the external environment. Sensory input from multiple modalities is routed back to the central nervous system and modulates motor commands. In parallel, mechanical preflexes directly act to resist perturbations. While this is visualized for insects here, it represents a general model for locomotor control and such structures are shared with other invertebrates and mammals (for more details see (Dickinson *et al.*, 2000)). This is visualized in c) in a schematic of differential roles of four components that underly locomotion across animals (following and adapted from (Ijspeert, 2018)). The schematic is not meant as a quantitative characterization, but should point out the shared and common characteristics between animal species. Complexity is as well only a rough ordering of animal species, it could be related to the number of neurons in the respective nervous systems (for more details see (Ijspeert, 2018)).

Dickinson (Dickinson *et al.*, 2000) pointed out that such properties of embodiment are an important part of adaptive behavior and interacting with an environment for all animals (Figure 2.1 b) shows a sketch of his conceptualization and c) adds an overview by (Ijspeert, 2018) on contributions of different factors across a spectrum of animals). Exploiting mechanical properties of the body (e.g., muscles) and mechanical preflexes can facilitate fast running and can compensate for small disturbances (for another example see (McGeer, 1993), passive walkers). Including characteristics as elastic properties into robots has already shown to be advantageous (Kim & Wensing, 2017; Schmitz *et al.*, 2008) and this integration is recognized as one of the major challenges for more adaptive robots (Yang *et al.*, 2018).

2.2 Hierarchical Organization in Biological Motor Control

One advantage of legged locomotion is that it allows animals to deal with quite difficult and uneven terrain. But this requires an adaptive motor system and control

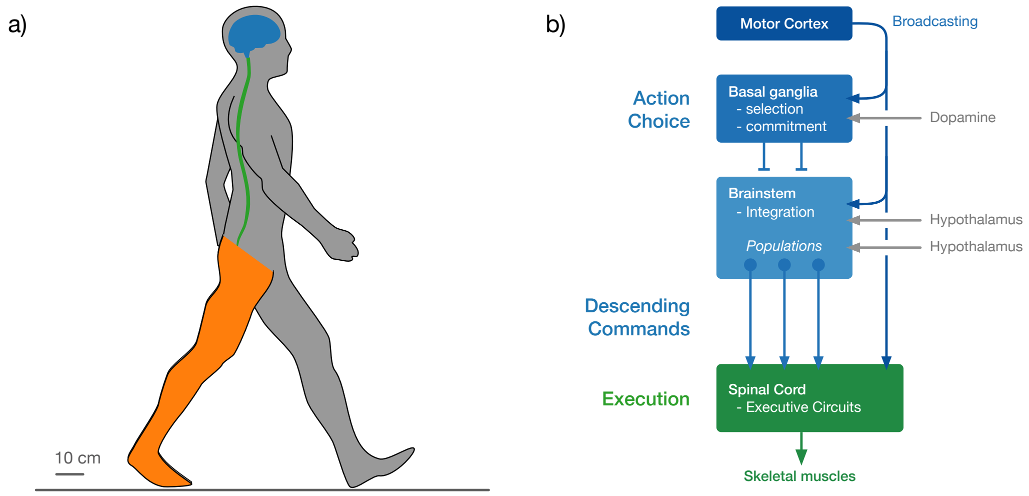


Figure 2.2: Overview of hierarchical control as assumed in humans: a) Sketch of motor control systems in human (corresponding to Fig. 2.1 a) for insects). Higher level control is color-coded in blue, central nervous system processing in green, and lower-level (sensory systems and embodied actuation) in orange. b) Schematic for hierarchical motor control and circuits for body movements in humans (adapted from (Arber & Costa, 2018)): Movements require coordinated activation of different neuronal populations across different parts of the nervous system. Higher centers project onto sensorimotor cortex which broadcasts to basal nuclei, brain stem, and spinal cord. On a high level (blue), an action is selected. These circuits transmit information concurrently for movement control to brainstem command lines (Graziano, 2006). Descending command lines from the brainstem activate executive circuits in the central nervous system (shown in green), e.g., for the control of high-speed locomotion or forelimb movement. On the lowest level (shown in orange), executive circuits control actuators that govern body movements. While this schematic is shown here for humans at the one end of the complexity spectrum, it is known as a general control strategy for vertebrate locomotion (Grillner, 2003) and the overall structure is assumed for other animals as well (Dickinson *et al.*, 2000). One difference between the hierarchical organization schematics for humans shown here in b) and the schematic shown for insects in 2.1 b) is the different and complementary focus. While research on vertebrates and humans puts a focus on higher level processing in the brain and descending commands, this is nicely complemented by research on embodiment and the contribution of mechanical properties in insects (see 2.1). The importance of these contributions is in general acknowledged for all animals (Arber & Costa, 2018; Graziano, 2006; Dickinson *et al.*, 2000).

structures that allow to deal with unpredictable environments, for example, during climbing or fast running. Such control structures have to coordinate fast movements of the animal as it is interacting with the environment. None-the-less, locomotion is a widespread trait that can be found across different animals, ranging from mammals to simple systems as are insects (Dickinson *et al.*, 2000).

The difficult task of controlling a complex system is addressed in many animals (including insects) through hierarchical organization and modularization of the control system in which the complexity is distributed onto different levels of a motor hierarchy (Botvinick, 2008; d'Avella *et al.*, 2015) and split into functional modules (Alon, 2006) (see Fig. 2.2). A highest level deals with selecting goal-directed behaviors. In our case, we are dealing with walking which is assumed an automatic behavior that is not planned in detail on a higher level and doesn't require attention. On an intermediate level, actions are selected (Arber & Costa, 2018) depending on context (Fig. 2.2 b) shown in blue). This leads to an internal competition between different actions that is context dependent.

The lower level realizes motor control primitives (or synergies) (Giszter *et al.*, 1993; Hart & Giszter, 2010) that are modulated by the higher levels through descending commands (Fig. 2.2 b), lower level shown in green with the higher levels' projections shown as descending commands). The lower level motor primitives allow for fast, sensory-guided adaptation towards disturbances. These describe how different muscles are working in concert for performing a specific action. Importantly, as a result of this distribution of complexity, the lower level is focused on small groups of muscles which leads to a decentralized organization and concurrent operation of multiple such lower level motor primitives. Such a modularization can be found in vertebrates and invertebrates (Flash & Hochner, 2005; Pearson, 1995). In vertebrates and higher animals the research focus is usually on the higher levels (Fig. 2.2 b) provides a good summary of such a state-of-the-art view (Arber & Costa, 2018), and similar ones can be found, e.g., in textbooks as (Magill & Anderson, 2017); higher levels shown in blue). Such a high level view is nicely complemented by work in invertebrates (Fig. 2.1) that demonstrates the importance of embodiment and mechanical properties (shown in orange in Fig. 2.2 a) and Fig. 2.1) as well as the importance of the lower-level motor primitives that are modulated by sensory inputs and descending commands from higher levels (shown in green in Fig. 2.2 and Fig. 2.1).

2.3 Hierarchical Organization in Technical Approaches

A hierarchical organization of motor control has been transferred to many robot control architectures. In general, this is realized as a distinction between selection of actions and execution of actions on two different levels of such a control hierarchy. For example, impressive work on dealing with walking on rough terrain comes from the area of quadruped robots (Carlo *et al.*, 2018). There, the problem is divided onto different control levels of a hierarchy. While on a lower level the detailed movements of joints and motors have to be controlled, a higher level coordinates movements between different legs and mainly selects lower level control primitives.

In many cases of locomotion control, on a higher level fixed gait patterns are assumed (Ijspeert, 2008). As one example, Kalakrishnan *et al.* (2010) used the LittleDog robot by Boston Dynamics to deal with challenging terrain. Spatial coordination is realized as a search for footholds: First, possible footholds are identified and a rough path is planned using a pre-trained ranking function and a scanned three dimensional terrain map. During locomotion, the detailed path is executed depending on the current posture as well as the preplanned schedule. This approach produced quite stable walking over rough terrain. A similar approach has been applied recently by Bellicoso *et al.* (2018) on the robot ANYmal. But importantly, they argue that different environmental situations also affect temporal coordination of legs on a higher control level. Therefore, they introduce a gait switching module that plans how to switch phases between fixed gaits.

2.4 Decentralization in Biological Motor Control

In contrast, behavioral results from insect walking studies show a wide diversity of walking behavior and not just a small number of fixed gait patterns (DeAngelis *et al.*, 2019; Bidaye *et al.*, 2018). Temporal coordination of locomotion appears better characterized as free gaits in which temporal relations emerge from the interaction with the environment. This allows to constantly adapt locomotion to unpredictable environments and to adjust the temporal coordination as required. A decentralized control structure appears to be crucial and beneficial for adaptivity of walking. Such an organizational structure of motor control in insects is well described and characterized by decentralization (Dürr *et al.*, 2004; Bidaye *et al.*, 2018). On the one hand, this agrees with a hierarchical organization as there is a higher level producing decisions which behaviors to perform. On the other hand, these findings further point out that biological control acts in a concurrent and modular fashion. Motor control for walking in insects is assumed to be constituted of local control modules. There exists one individual controller for each leg that switches local behaviors depending on current sensory signals (Schilling *et al.*, 2013*b*). These controllers coordinate their behavior through, on the one hand, local coordination rules that influence the switching behavior (see Fig. 2.3 b). On the other hand, the different controllers are coupled through the body and the interactions with the environment. Actions of one leg affect other legs and can be sensed by those without requiring explicit information exchange (at least to a certain degree). In this way, the system exploits the loop through the environment (Brooks, 1991).

Decentralization constitutes a key characteristic of motor control and is one focus of this thesis: The first article (Schilling *et al.*, 2013*a*, see article A.1 in the appendix, page 94) provides a detailed overview on behavioral findings in stick insects and introduces the decentralized control architecture Walknet for six-legged walking in which each leg is controlled concurrently by an individual controller. In (Schilling & Cruse, 2020, see article A.2, page 118) complementing neuroscientific results are presented and the control architecture is realized using a more detailed and biologically realistic neuron-type, showing temporal characteristics and addressing control on the joint level.

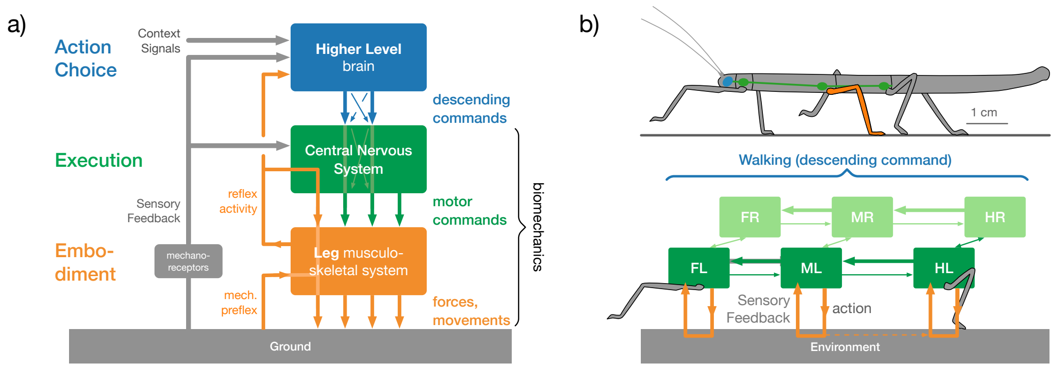


Figure 2.3: Decentralization as an important characteristic for motor control: a) provides an overview schematic of our motor control perspective. First, motor control is hierarchically organized, integrating different levels of neural organization as assumed in animals (as detailed in Fig. 2.2) and complementing work on, for example, insects highlights the importance of musculo-skeletal properties as well as the interaction with the environment (as detailed in Fig. 2.1). Second, there is a decentralized organization operating on different timescales. Behavior emerges as a result of decentralized and locally interacting concurrent control structures. Note, that there are multiple, parallel arrows connecting different levels. Such a concurrency is well established in insects (Schilling *et al.*, 2013*a*), but is also present in vertebrates (Graziano, 2006, see Fig. 2.2 b)) which, unfortunately, is only rarely explicitly pointed out. In b), this is shown for the decentralized Walknet system for the control of six-legged walking which reflects a control architecture as found in stick insects (Schilling *et al.*, 2013*a*).

2.5 Walknet – a Decentralized Control Architecture

Decentralization and hierarchical organization have been realized as control principles in the Walknet system that has been applied in simulation and on different robots (Dürr *et al.*, 2019; Schilling *et al.*, 2013*a*). While Walknet is structurally quite a simple system it is adaptive and can deal with severe disturbances as for instance loss of a leg. In Walknet, control is distributed hierarchically onto different levels. Each leg has its own controller (Fig. 2.4 shows arrangement of six such leg controllers) that locally decides which action to perform depending on the sensed context (Fig. 2.5). For locomotion, it is distinguished between two basic actions on a leg level, protraction and retraction (stance and swing movement, respectively). In forward walking, switching from a swing movement to the front towards a stance action is initiated after the leg touches the ground and starts carrying weight. During stance mode, the leg contributes to carrying the body and propels the body forward. The transition from stance to swing is determined by the position of the leg. Swing is started when the leg moves behind a posterior extreme position (PEP).

Each leg controller consists of competing motivation units (Fig. 2.5) that are associated with the different possible behaviors. Each action is connected with one such unit that represents the activation of that action and the units form a local winner-take-all network that decides which action is active. As mentioned, action

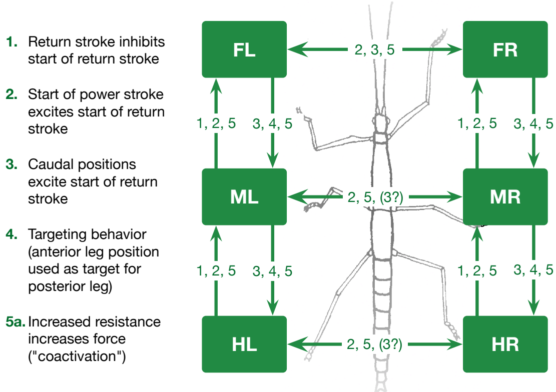


Figure 2.4: Schema of the morphological arrangement of the leg controllers and the coordination influences (1–6) between legs (adapted from (Schilling *et al.*, 2013*b*)). Legs are marked by L for left legs and R for right legs and numbered from 1 to 3 for front, middle, and hind legs, respectively. The question mark indicates that there are ambiguous data concerning these influences.

selection is mostly sensory-driven and depends on the current state of the controller and of that particular leg. Furthermore, the six leg controllers coordinate their action, but rely on local information: Neighboring legs are connected through local coordination influences (Fig. 2.4) that are derived from behavioral experiments on stick insects. Legs influence when neighboring legs starts to produce swing movements by translation of the PEP. This local coordination is sufficient to prevent that, for example, two neighboring legs are lifted from the ground at the same time.

Details on the Walknet control approach are given in the first article in the appendix (Schilling *et al.*, 2013*a*, A.1, page 94) which also provides a review on behavioral experiments from which the coordination influences were derived. The Motivation Unit architecture is provided in a further publication (Schilling *et al.*, 2013*b*, see A.3, page 168), explaining the network organization in detail and showing results for application on a hexapod robot. Finally, this has been further extended towards a detailed decentralized architecture that acts on the joint level in neuroWalknet (Schilling & Cruse, 2020, see article A.2, page 118).

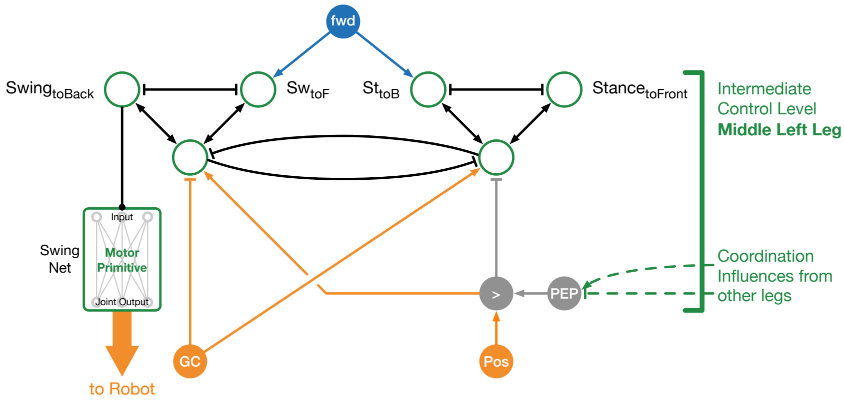


Figure 2.5: Local leg controller (shown for a single leg). Green units represent the intermediate local leg level on which actions are selected. These units are called motivation units and each one is connected to a motor primitive which is modulated by the activation of the motivation unit that drives the behavior (Sw_{toF} is a swing movement directed towards the front, St_{toB} a stance movement that is moving the leg to the back). Competition between behaviors (shown are swing and stance in two directions) is realized through mutual inhibition. Global information is indicated again in blue (in this case only walking direction is provided as global information). Sensory information (orange) drives switching of behaviors in context-dependent manner. On the one hand, detected ground contact (GC) inhibits swing movements and starts stance. On the other hand, the leg position (Pos) is compared to an assumed posterior extreme position (PEP) which determines when to initiate a swing movement. Coordination influences modulate the PEP, facilitating early swing movements or prolonging stance phase.

2.6 Results Overview

Results from the three publications on local control can be summarized as, first, free gaits emerge from the decentralized control structure which allows to continuously adapt on a short timescale to changing environmental conditions. Second, on a neural level, activation of neurons reproduces experimental findings from insects. These characteristics are visualized in Fig. 2.6 to 2.8.

First, a continuum of different free gaits emerges (Schilling *et al.*, 2013*a*)—depending on the velocity of walking—from the interaction of the concurrent local control modules (Fig. 2.6). Importantly, such a system is quite adaptive: Such temporal coordination patterns converge towards stable gaits in undisturbed environmental setting, but excel in difficult and disturbed experimental scenarios (Schilling *et al.*, 2013*b*). When dealing with disturbances the temporal coordination emerges directly from the interaction between the concurrent controllers as needed. For example, in curve walking there are no clear-cut phase relations between different legs (Fig. 2.7) which poses a problem for centralized control approaches (see (Dürr *et al.*, 2019)) or in the case of climbing, adjusting and searching for footholds requires time which directly affects temporal coordination between different legs. Insects adapt to these challenges and show flexible walking behavior. Therefore, such a decentralized control structure appears beneficial as well for robots which

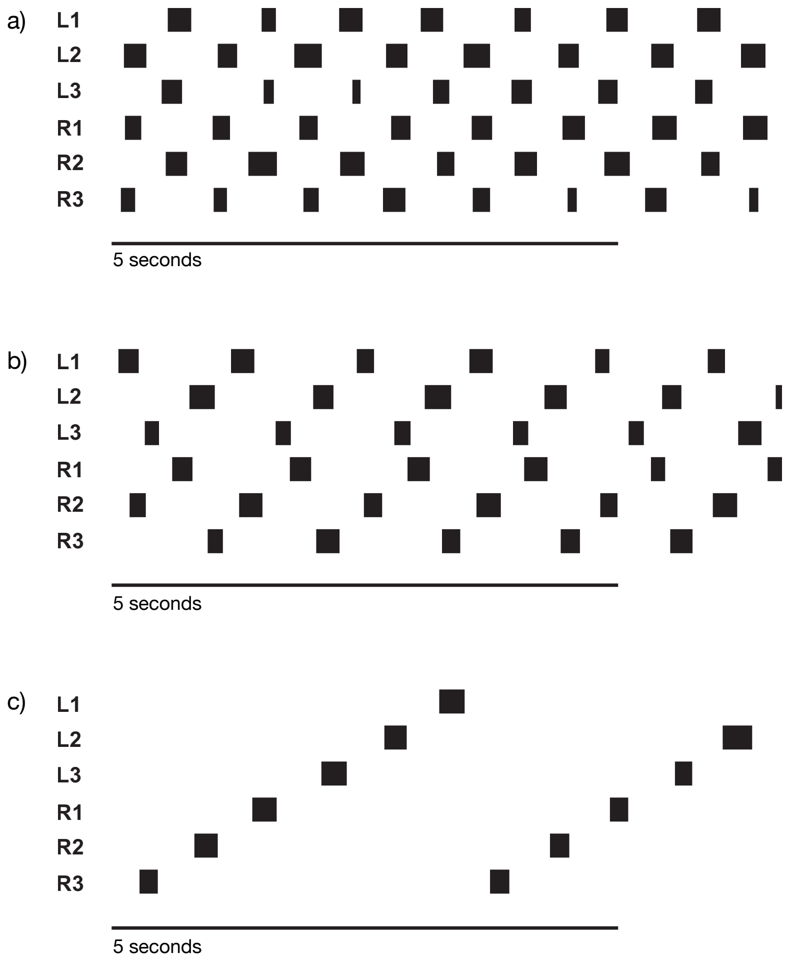


Figure 2.6: Simulated robot Hector walking straight for different velocities (Schilling *et al.*, 2013*b*): a) high velocity leading to emergence of a tripod gait, b) moderate velocity leading to tetrapod gait, and c) low velocity leading to a wave gate pattern. Black bars indicate swing movement of the respective leg: left front, middle and hind leg, right front, middle and hind leg, from top to bottom. Abscissa is simulation time. The lower horizontal bars indicate 500 iterations corresponding to 5 s real time.

has been demonstrated in numerous dynamic simulations and on real robots as well (Dürr *et al.*, 2019; Schmitz *et al.*, 2008).

Second, the proposed control structure allows to analyze neural activity inside the controller in detail. Intrinsic rhythmic behavior is often assumed to be caused by Central Pattern Generators (CPG) (Ijspeert, 2008; Orlovsky *et al.*, 1999). The recognition of CPGs as a basic building block in control is based on experimental findings in studies of deafferented animals (Pearson, 1995; Orlovsky *et al.*, 1999). In deafferentation, sensory input and motor output is operationally interrupted. During the experiment the neural system is artificially stimulated by, for example, application of pilocarpine (Büschges *et al.*, 1995). In such experiments, observed temporal coordination in the deafferented animals shows an in-phase coupling between neighboring legs. While such an in-phase coordination between contralateral

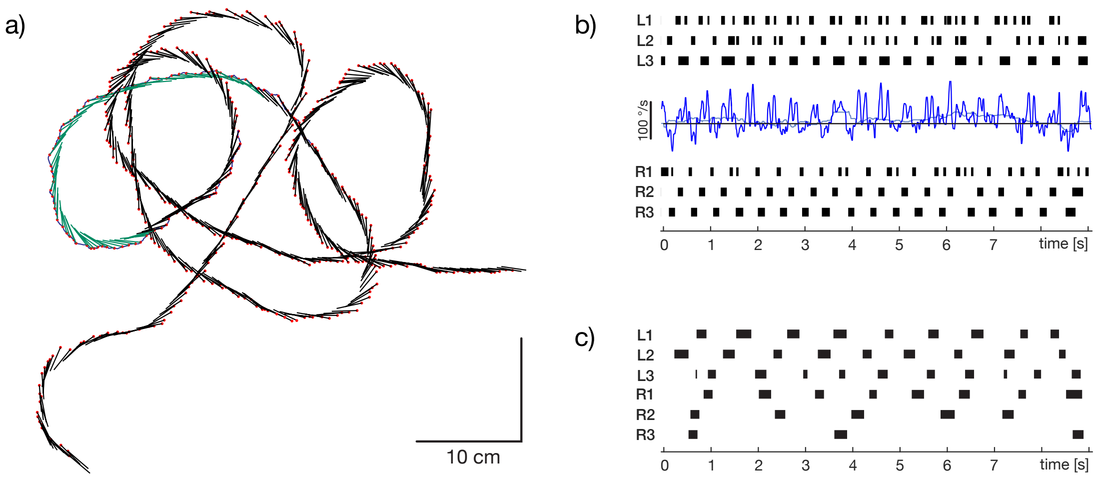


Figure 2.7: Example of curve walking in stick insects. a) Shows a sequence of a free walking, blindfolded stick insect on a horizontal plane (Dürr *et al.*, 2019). Black line segments and red dots show body axis and head every 200 ms (overall duration: 106 s; median speed was 35 mm s^{-1} at the beginning (lower left) and 25 mm s^{-1} at the end). Green line highlights the part shown in the footfall pattern shown in b) on the right.
b) Footfall pattern with black lines showing swing episodes of all six legs (L1 to L3: left front to hind legs; R1 to R3: right front to hind legs) and corresponding yaw rotation of the body axis. Blue lines show median rotational velocity per 60 ms window (thin dark blue) and per 1 s window (thick light blue). c) Footfall pattern for simulated robot Hector walking a turn to the right (Schilling *et al.*, 2013*b*). The complete run shown corresponds to a turn of about 180°. Starting positions (in m, origin is position of coxa): L1: 0.20, R1: 0.05, L2: -0.04, R2: -0.14, L3: -0.02, R3: -0.22.

legs may corresponds to behavior for the case of swimming or flying, it doesn't match to walking behavior which is characterized mostly by anti-phase coupling of neighboring legs (Graham, 1972; Wosnitza *et al.*, 2013). None-the-less, CPGs are often generalized to control locomotion even though the coordination of deafferented leg controllers is quite different, eventually even opposite to that of normal walking. The decentralized Walknet approach challenges such a CPG-based perspective as in Walknet action selection is not driven by rhythmic oscillations, but sensory driven (Schilling *et al.*, 2013*a*). Whereas the original Walknet couldn't account for findings of neural in-phase oscillations—as it operated on a more abstract intermediate level of action selection and used a simpler neuron model—, in an extension, neuroWalknet now introduced a detailed neuron model that shows temporal behavior on the neuronal level and can reproduce these findings (Schilling & Cruse, 2020). Note, this is not meant to neglect the general existence of slow CPGs to control rhythmic behavior. We only show that a sensory-driven control approach can show similar activation patterns without the need of explicit CPG structures.

As one example, the simulations in Fig. 2.8 show that results of Knebel *et al.* (2017) can be reproduced. In experiments on deafferented locusts, these authors found in-phase coupling between neighboring legs if all three thoracic ganglia were treated with pilocarpine, which again contrasts to normal walking behavior, but

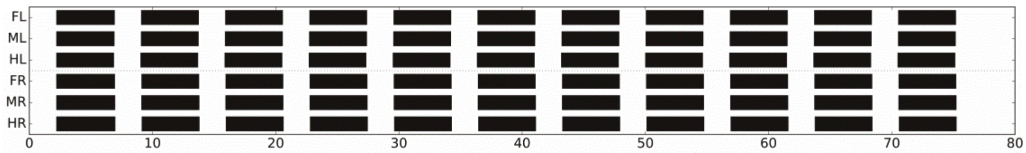


Figure 2.8: Neural activity of depressor activation over time (s) during chemical activation (Schilling & Cruse, 2020): Shown is a simulation of experiments of deafferented locusts (Knebel *et al.*, 2017). Black bars show activation of depressor muscle output ($> 0\,\mathrm{mV}$). The main results are that all six hemiganglia oscillate with a period of about 5 s and show in-phase coupling (recordings from depressor motor neurons), if all three thoracic ganglia were treated with pilocarpine. Note that influences from rules 1-3 were not effective in this situation as sensory input concerning leg position is fixed (and leg controllers were assumed to be in stance mode during the application of pilocarpine). Therefore, the critical effects resulted from rule 5 influences plus the extension assuming a contralateral inhibitory connection between the hind leg controllers.

can be observed in swimming (Ikeda & Wiersma, 1964) and flying (Pearson, 1995). Such a rhythmic neural activity could be reproduced in neuroWalknet even though the leg controller does not contain any explicit CPG that triggers the rhythmic movement during normal walking, i.e., controls patterns characterized as pentapod, tetrapod or tripod (Schilling & Cruse, 2020). When all three ganglia in the different body segments were treated with pilocarpine, all six hemiganglia oscillated in-phase. These results raise questions concerning the contribution of the role of oscillatory systems in the control of walking.

2.7 Conclusions

This chapter highlighted complementing findings from biology that provides insights into the structure of control systems in animals and showed how these can be applied for motor control on robots (for a schematic see Fig. 2.3).

1. **Embodiment:** Adaptive behavior exploits mechanic properties as are elasticities of muscles. Furthermore, kinematic interaction with the environment can simplify and replace the need of costly internal computation and representation.

2. **Hierarchical Organization:** In animals, motor control problems are distributed onto different levels and into different modules. There is ongoing competition within these levels, for example, in action selection. Further, there is interaction between the different modules, for example, higher levels modulate lower levels through descending commands. This allows to focus on specific contexts and situations as well as dealing with different temporal scales. While the higher level deals with goal selection, lower levels deal with sequential action selection and on the even lower levels with control of muscles.

3. **Decentralization and Concurrent Processing:** Motor control on the lower level is not only driven by descending commands from higher levels. A lower level is constituted by local, decentralized control circuits or motor primitives that allow for fast, sensory-guided adaptation towards disturbances. This complements the higher levels that deal with action selection on a longer timescale and modulate the lower levels. Importantly, these decentralized control structures process information concurrently.

These control characteristics have been applied on a hexapod walking robot in our Walknet control approach. As each leg is controlled by an individual, decentralized controller, overall behavior emerges from the interaction of these concurrent and hierarchically organized controllers. Coordination between controllers is realized through, on the one hand, local coordination influences acting only between neighboring legs. On the other hand, there is an implicit, embodied coordination as effects of the other legs are mediated through the loop through the world (Brooks, 1989). This control structure produces quite adaptive behavior and can deal with disturbances acting on different timescales (which will be discussed in the following chapter), for example, dealing with quite severe interventions after the loss of a leg or when it becomes necessary to continuously adapt as during climbing through a twig. But as one disadvantage, the resulting control structures were handcrafted, which requires a high level of expertise (as can be seen for the detailed schematic of the neuroWalknet (Schilling & Cruse, 2020)). Scaling such designed approaches further to more varied real world environments appears difficult due to an increasing number of required neural units and sensory inputs that have to be integrated and the possible interactions between all the different concurrent control parts. It appears better to turn towards a learning approach that allows to self-improve over time, as will be introduced in the next chapter.

2.8 List of Publications

This chapter gave a brief introduction and summary of three publications that are part of this thesis and can be found in the appendix.

2.8.1 Contributions to the Thesis

- Schilling, M., Hoinville, T., Schmitz, J. and Cruse, H. (2013), "Walknet, a bio-inspired controller for hexapod walking". Biological Cybernetics, 107(4), pages 397–419.

  Appendix A.1, page 94: Provides a review on behavioral findings in insects and details on the Walknet control approach (Schilling *et al.*, 2013*a*, A.1, page 94). This was published in Biological Cybernetics (Impact Factor 1.76) and is now established as a reference for decentralized organization of motor control in insects (cited 129 times).
  Author Contributions: MS and HC laid out the concept, designed the model and analyzed the data. MS carried out the implementation. MS and HC wrote the manuscript. Writing review and editing all authors.

- Schilling, M. and Cruse, H. (2020), “Decentralized control of insect walking - a simple neural network explains a wide range of behavioral and neurophysiological results”. PLOS Computational Biology 16(4): e1007804. https://doi.org/10.1371/journal.pcbi.1007804

  Appendix A.2, page 118: Extension towards a detailed decentralized architecture that acts on the joint level. This was recently accepted for publication in PLOS Computational Biology (Impact Factor 4.43).
  Author Contributions: Conceptualization, methodology, investigation, and writing – MS and HC. Software, simulation, and data curation – MS. Formal analysis – HC.

- Schilling, M., Paskarbeit, J., Hüffmeier, A., Schneider, A., Schmitz, J., and Cruse, H. (2013), “A hexapod walker using a heterarchical architecture for action selection”. Frontiers in Computational Neuroscience 7:126. doi: 10.3389/fncom.2013.00126.

  Appendix A.3, page 168: Introduces the hierarchical Motivation Unit architecture and shows results for application on a hexapod robot. The article was published in Frontiers in Computational Neuroscience (Impact Factor 3.57).
  Author Contributions: Conceptualization and writing – MS and HC. Methodology and investigation – MS, HC, JS, and AS. Software and simulation – JP, AH, and MS.

2.8.2 Further Related Publications

The summarized characteristics and findings have lead to further publication including, for example, application on the real robot Hector.

Peer-reviewed Journal Papers:

- Dürr, V., Arena, P., Cruse, H., Dallmann, C.J., Drimus, A., Hoinville, T., Krause, T., Mátéfi-Tempfli, S., Paskarbeit, J., Patanè, L., Schäffersmann, M., Schilling, M., Schmitz, J., Strauss, R., Theunissen, L., Vitanza, A., and Schneider, A. (2019), “Integrative Biomimetics of Autonomous Hexapedal Locomotion”. Frontiers in Neurorobotics 13:88.

Book Chapters:

- Cruse, H. and Schilling, M. (2018), “Pattern Generation”. In: Living Machines: A handbook of research in biomimetic and biohybrid systems. T. J. Prescot, N.F. Lepora, P.F.M.J. Verschure, (eds). Oxford University Press, p. 218-226.

- Schack, T., Bläsing, B., Hughes, C., Flash, T. and Schilling, M. (2014), “Elements and Construction of Motor Control”. In A. Papaioannou and D. Hackfort (Eds.), Routledge Companion to Sport and Exercise Psychology:Global Perspectives and Fundamental Concepts. London: Routledge (pages 306–321).

Reviewed Conference Proceedings:

- Hoinville, T., Schilling, M., and Cruse, H. (2015). “Control of rhythmic behavior: Central and Peripheral Influences to Pattern Generation”. in: ICRA Workshop on ‘Pros and Cons of Central Pattern Generators’, ICRA 2015, pp. 1-3.

- Schneider, A., Paskarbeit, J., Schilling, M. and Schmitz, J. (2014), “HECTOR, A Bio-Inspired and Compliant Hexapod Robot”. In: A. Duff, T. Prescott, P. Verschure, N. Lepora (eds.): Living Machines 2014, LNAI 8608, pp. 427–429.

- Cruse, H., Schilling, M. (2014). “Action Selection within short time windows”. In: A. Duff, T. Prescott, P. Verschure, N. Lepora (eds.): Living Machines 2014, LNAI 8608, pp. 47–58.

Learning from scratch can be overwhelming, as it involves relations between motor and perceptual skills, resulting in an extremely large dimension search problem.

— Montesano *et al.* (2008, p. 16)

3
Learning of Adaptive Behavior in a Decentralized Fashion

Contents

3.1 Hierarchical Representation in Robotics 30
3.2 The Advantage of Decentralized Control 32
3.3 Decentralized Deep Reinforcement Learning 36
3.3.1 Reinforcement Learning 36
3.3.2 Decentralized Motor Control Architecture 37
3.3.3 A Centralized Approach as a Baseline 38
3.3.4 Policy Network 38
3.4 Results Overview 38
3.5 Conclusions 40
3.6 List of Publications 42
3.6.1 Contributions to the Thesis 42
3.6.2 Further Related Publications 43

In the previous chapter, hierarchical organization and decentralization were introduced as biological characteristics of motor control and were applied in the Walknet structure for six-legged walking. We have seen how free gaits emerge in these local control systems depending on the velocity of walking and the current environmental settings. Such temporal coordination patterns converge towards stable gaits in undisturbed environmental settings, but the system excels in difficult and disturbed experimental scenarios. In such cases, for example, consider the loss of a leg or constant adaptation of footholds as required during climbing, such controllers produce walking behavior by continuously adapting to the changing environmental conditions which has shown to produce robust and stable behaviors. But in this approach—as well as in many other biological inspired approaches—control structures are often handcrafted. While this allowed to show how such systems can deal with—often quite challenging—disturbances, these systems are tested only in a small number of quite specific experiments that are lacking the general adaptivity as found in animals. Scaling such designed approaches to real world environments simply appears impossible due to the increasing number of required sensory inputs that have to be integrated and the increasing number of possible interactions between different concurrent control parts (Aljalbout *et al.*, 2020). Therefore, such approaches are usually not tested for generalization, for example, when climbing in difficult terrain. Here, learning-based approaches offer the advantage of avoiding handcrafted modules, but rather the details of the control structure evolve over time.

In this chapter, we will use a Deep Reinforcement Learning (DRL) approach on an embodied—currently simulated—robot. As a scientific question, we are interested in how the biological characteristic of decentralization affects, on the one hand, the learning process, and, on the other hand, performance of learned control structures. Does a decentralized control structure allow to learn usable control structures and how does the performance of these learned controllers compare to a baseline approach? Learning of the parameters of the decentralized control—as well as the baseline—structure (the policies) is reward driven from interaction with different environmental conditions. We hypothesize that a decentralized architecture of six local leg controllers—that each control the behavior of a single leg relying only on local information—is sufficient to produce adaptive locomotion behavior in difficult and changing environmental settings. This is tested through comparing learned decentralized controllers with a centralized approach as a baseline on a six-legged robot and shows that the decentralized approach performs at least as well, if not even better. The chapter is organized in the following way: first, we briefly give an example for the current state-of-the-art in DRL of locomotion. Current DRL approaches already advocate hierarchical organization (Merel *et al.*, 2019*a*), but in the second section we will argue why such a representational approach to hierarchical organization appears not sufficient to explain adaptive walking behavior, but only can deal with—however quite severe—disturbances on a long timescale. Third, the decentralized learning system is introduced followed by a summary of the results. Last, the conclusion gives an outlook on learning in a hierarchical and decentralized control structure.

3.1 Towards Learning Hierarchical Representation in Current Robotic Approaches

Learning based approaches offer the advantage of not requiring handcrafted modules. Recently, Deep Reinforcement Learning (DRL) was established as one promising approach (Arulkumaran *et al.*, 2017). But while it has shown to become more sample-efficient and stable with respect to variability of experimental settings during training (Fujimoto *et al.*, 2018), high dimensionality of a control problem still imposes a difficult problem. Even more when turning towards application on

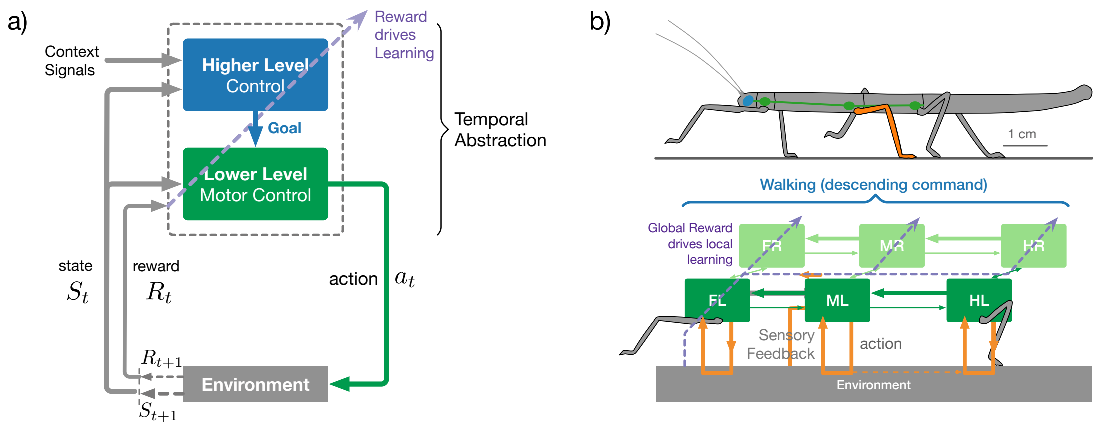


Figure 3.1: Visualization of influences for the biological inspired approach: On the left (a) the standard view of interaction with the environment in reinforcement learning (Sutton & Barto, 2018) is extended to a hierarchical perspective (Kulkarni *et al.*, 2016) as advocated, for example, in (Merel *et al.*, 2019*a*). For higher level control (shown in blue) this is in agreement with what we know on the structure of motor control in mammals (Arber & Costa, 2018) about descending pathways and modulation of lower level control centers (shown in green) in the spinal cord. Such structures are shared not only in mammals, but also in invertebrates and insects (Dickinson *et al.*, 2000), see b). Work in such simpler model systems allows a more detailed analysis of interaction with the environment which has stressed the importance of very fast and local reflex activity controlled directly on the lowest level or that are even realized by passive properties as muscle elasticities or preflexes (shown in orange). One important characteristic emphasized by this work is the emergence of behavior as a result of decentralized and locally interacting concurrent control structures (bottom part of b). An example is given by the decentralized control structure found in stick insects (Schilling *et al.*, 2013*a*), but this concurrency is as well assumed in primates (Graziano, 2006).

robots. In that case, learning appears difficult or simply infeasible if no additional structure for the control system is given (Hwangbo *et al.*, 2019). Therefore, a lot of this work has only been realized in simulation or uses at first simulation before transferring a learned controller onto a real robot. For example, Hwangbo *et al.* (2019) used DRL for, first, training a policy network in simulation which, in a second step, was successfully transferred to the real robot system. As one limitation the authors point out that the learned behaviors tended to overfit (Lanctot *et al.*, 2017) and did not show adaptivity as found in animal systems. As a possible solution the authors proposed that a hierarchical organization of the controller might help to alleviate this problem.

Hierarchical Deep Reinforcement Learning (HDRL) has become more and more prominent (Merel *et al.*, 2019*a*; Arulkumaran *et al.*, 2017) and has been applied to locomotion in simulation as well (Frans *et al.*, 2018; Heess *et al.*, 2016). HDRL (see Fig. 3.1) provides a biologically inspired solution which can be traced back to the early options framework which is described on different time scales realizing a form of temporal abstraction and hierarchical representation (Sutton *et al.*, 1999).

Current approaches turn towards how to self-discover options on the higher level (Kulkarni *et al.*, 2016; Vezhnevets *et al.*, 2017). This has lead to some early results and the introduction of more stable learning approaches (Fujimoto *et al.*, 2018; Nachum *et al.*, 2018) is assumed to further leverage such approaches towards more and more real world problems. But most of these approaches maintain the notion of fixed timing patterns between the different levels (as one example for an exception see (Han *et al.*, 2020)), i.e., the higher level policy is only evaluated after a fixed number of steps of the lower level to enforce a form of behavioral stability. This works well in some scenarios, but when turning towards scenarios in which there is a large variation concerning the duration of different behaviors or actions, this fixed stepping on the different levels might cause problems.

To summarize, current Deep Reinforcement Learning approaches start to incorporate a hierarchical organization. This shows to be beneficial as such approaches can deal with a wider variety of contexts. Hierarchical organization for locomotion control has already been tested in simulation and is seen as a promising approach for real robot application. But mostly this success is connected to transfer learning and the ability to switch between different contexts which allows to exploit a control structure of a similar context. This has shown to work well in the above examples were different low-level motor primitives are used for walking behaviors and are combined on a higher level for navigation. Such approaches also tend to work well when dealing with severe intervention, for example, the loss of a leg which is often used as an application scenario. But such an intervention presents a singular event which might require adaptation of the control structure (our Walknet approach can deal with loss of leg without introducing specific structures (Schilling *et al.*, 2007)). This is in contrast to adaptivity as found in insects, for example, considering climbing through a twig or even simply negotiating a curve. In this case, behavior is continuously adapted to the unpredictable environment as it is sensed in real-time. This appears to require other mechanisms as well that act on a much faster timescale. Along this line, we argue that decentralization is an important characteristic and we believe that a dynamical systems perspective of temporal coordination—as introduced in the previous chapter—illustrates this point nicely. This view has recently been supported by Peng *et al.* (2019) who provided a similar argument in their hierarchical approach of learning compositionable (concurrently executed) skills that allowed for better transfer between tasks.

3.2 The Advantage of Decentralized Control

In this section, we want to motivate our argument for the importance of decentralization in motor control. From our point of view, current hierarchical approaches take a representational stance (for a detailed discussion on emergence of representation in neural networks see (Brette, 2019)), i.e., the hierarchical organization in the respective approaches is reflected as a hierarchical representation. Such an organization is tied towards specific contexts and problems. Using navigation in a maze as a task (see Fig. 3.2),Lehman & Stanley (2011) illustrated

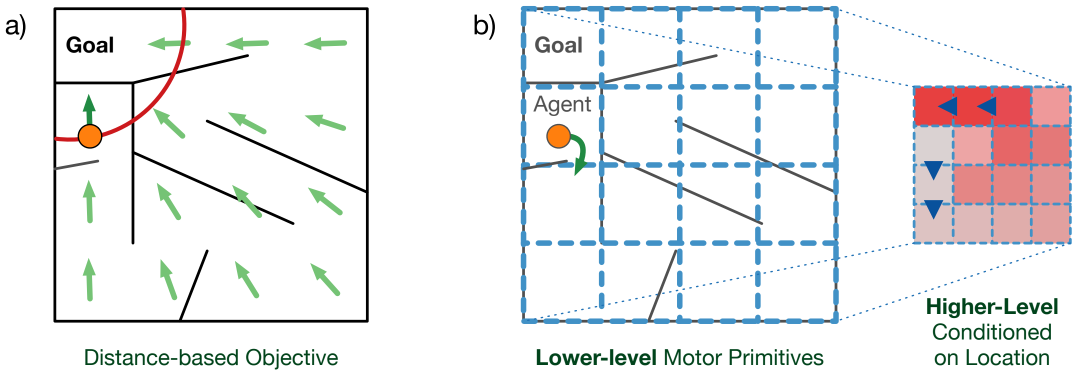


Figure 3.2: Deceptive Problems: a) Showing a simple maze navigation task (black bars represent obstacles/ walls inside the maze) with the goal at the top left (following Lehman & Stanley (2011)). When using a distance based objective function, all points on a circle are conflated and share the same reward value even though they clearly differ in how well they provide a path towards the goal location. This explains the idea of deceptive problems: the objective function is not informative for solving the problem as in many instances one has to move towards locations that appear worse before improving towards a global optimum. As a consequence, in such problems optimization often ends up in local minima. b) visualizes the introduction of a suitable hierarchical representation, in this case a grid-based map structure. Each square of this higher level structure is assigned a reward value based on experience (this still requires exploration) and transitions inside the grid are learned. During exploitation this map structure allows to follow the gradient along possible transitions even though this might result in moving away from the goal location at first as is shown for the example when recovering from a dead end.

nicely that a simple objective as distance towards the goal is not always helpful to bootstrap learning for finding a route towards a goal (Fig. 3.2 a)). Instead, they showed (Fig. 3.2 b)) that introducing a higher level map-like representation helps to solve the problem: On this higher level a grid like structure is mapped onto the original maze. Following a reinforcement learning-based approach values and actions are learned for each of these grid cells from real routes passing towards that subarea and integrating the accumulated rewards. This representation can afterwards be exploited to make higher level decisions for the agent as he is selecting actions leading into neighboring grid areas that have a higher expected reward. The associated lower level carries out the respective action. This has shown as a powerful approach for spatial problems, for example, see Kulkarni *et al.* (2016) on learning difficult delayed rewards in computer games through intrinsic motivation by using a hierarchical map-like representation. But, importantly, this requires finding such a suitable intermediate representational layer. Here, we want to argue that in many application areas it is not straight-forward to find such a representation. Ideally, such a representation is induced by the given task and can be derived from outside. In spatial cases, we can even observe activations that reflect such a spatial organization in the brain of animals (Moser *et al.*, 2008). But there is a different possible point of view, that there are no fixed representations, but that we merely observe an epiphenomenon while the regularity appears as an effect of

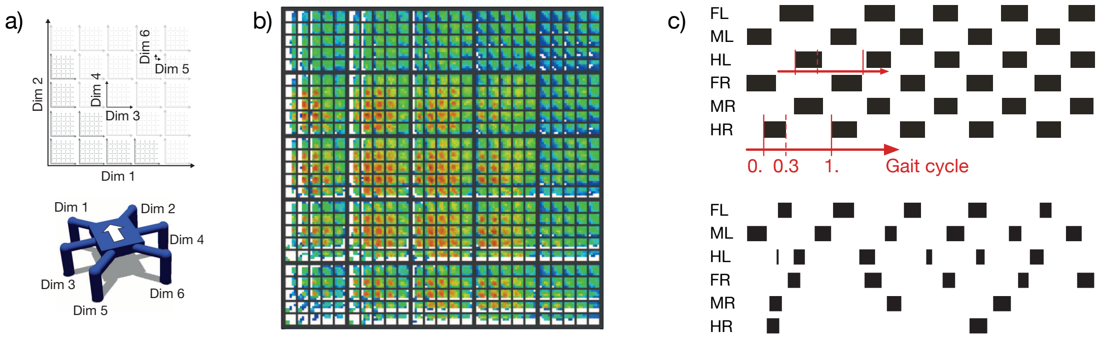


Figure 3.3: Visualization of behavioral space (a) and b) are reproduced from (Cully *et al.*, 2015)): a) Overall behavior is represented in an abstract, six-dimensional space (one dimension for each leg). Each dimension is the portion of time that each leg is in contact with the ground (see explanation of gait cycle visualized in red in the top of c)). Arrangements of the six dimensions in order to produce two dimensional color coded plots. The space is discretized at five values for each dimension. b) Behavioral map that stores one optimal behavior for each point in the six-dimensional behavioral space. Each colored pixel represents the highest performance for an associated behavior which were pretrained over the full six dimensional space for a simulated robot (shown in bottom of a)). c) shows, top part, a typical footfall pattern for a straight walking insect as produced by our Walknet architecture. In red it is shown for one leg (hind right leg) how the gait cycle is determined as the portion of the leg in swing compared to the complete stepping cycle. On the bottom, curve walking is shown as an example highlighting that in a realistic—still quite simple—context footfall patterns might be highly irregular and there appears neither a fixed stepping cycle nor a fixed length for swing or stance movements which makes this appear as an improbable space for motor control in such a setting.

the regularity of the context instead of being intrinsic or encoded in the neural structure (Such an epiphenomenal view is argued for by Brette (2019) for the case of spatial neural codes as found in place cells.).

To explain this in more detail, we will turn towards the impressive evolutionary approach by Cully *et al.* (2015) who argue that there is a behavioral space organizing behavior. They used an evolutionary algorithm to evolve locomotion behaviors for all kind of possible situations and build an exhaustive map of possible behaviors. Importantly, they used a representational approach, too. While the original learning problem is very high dimensional and appears intractable, they introduced an intermediate (lower dimensional) representational layer as a behavioral space. Their approach aims at finding a cover of optimal solutions across this behavioral space which is realized through a specific evolutionary algorithm that performs a broadened search (Mouret & Clune, 2015). The behavioral space describes temporal relations (in their case duty cycles of individual legs) and the authors were able to show that this intermediate representation could be used as stepping stones (Lehman & Stanley, 2011) for adaptive control. Following this approach, the robot was able to cope with leg damages. It could simply switch towards other behaviors using the previously derived behavioral space, for which they used a form of trial-and-error search across the low dimensional behavioral space. But while this has shown to work efficiently when experiencing drastic changes (as losing a leg), such a fixed temporal coordination scheme does not show the form of adaptivity we find in walking animals. The authors assume that animals have an understanding of the space of all possible behaviors. In Cully *et al.* (2015), central to their approach is how the behavioral space factorizes a current situation describing temporal relations. While temporal patterns are traditionally used in order to describe how animals move, we, in contrast, believe that this only provides a good observational space, but not one used as a substrate for motor control as indicated by the huge variability found in many studies in (slow) walking insects (Bidaye *et al.*, 2018; DeAngelis *et al.*, 2019; Schilling *et al.*, 2013*a*). Adaptive walking or climbing appears hard to be described by phase relations between different legs as those are changing all the time. Current approaches that employ fixed gait patterns run into problems when environmental conditions are changing (e.g., during acceleration or deceleration)—control approaches using fixed gaits actually produce poor results in these instances. The approach presented by Cully *et al.* (2015) appears well equipped to adjust leg patterns after a strong change, but this type of representation appears not well suited for moderating continuous change. As one example, we want to point out the temporal pattern observed when an insect negotiates a curve (see Fig. 3.3, for more details on curve walking in insects see Fig. 2.7): during curve walking, inner and outer legs as well as the different legs on one side of the animal, are walking with very different and constantly changing stepping frequencies as the legs are contributing differentially (e.g., inner hind legs are standing still in narrow turns and provide an anchor point). The underlying temporal relations emerge and while there is some structure to the phase relations, there is huge variability between trials and it appears implausible that a higher level control system is planning these temporal relations in detail (Schilling *et al.*, 2019).

Instead, it might help to reconsider the behavioral space of animals', not taking it as one specific type of static representation, but more as an observation of invariances that stem from the underlying—lower level—systems. Instead of enforcing a particular representation scheme, the interaction between the different (simple) dynamical systems span a behavioral space that can be exploited for adaptive behavior in different ways. It might, therefore, be a better approach to understand how the behavior emerges from interactions of different decentralized and concurrent low-level control modules as proposed in our system (Shenoy *et al.* (2013) argued for such a shift towards a dynamical system perspective to better understand the neural basis for grasping.). While such concurrent control circuits and their dynamic interactions appear more difficult to analyze and understand, this might be a better starting point to approach adaptivity and might be required to allow for a flexible hierarchical organization. Such a modular organization of the nervous system of animals has been linked to adaptivity (Kashtan & Alon, 2005; Lipson *et al.*, 2002).

This argument is discussed in more detail in the appendix (Schilling *et al.*, 2019, B.1, page 188).

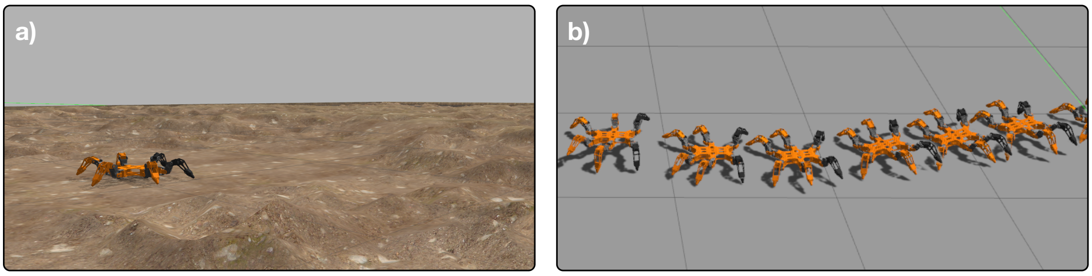


Figure 3.4: Visualization of the simulated PhantomX robot. a) shows the uneven terrain condition in comparison to the robot. b) shows a walking sequence on flat terrain.

3.3 Decentralized Deep Reinforcement Learning

In the previous chapter, we presented a decentralized dynamical systems approach. Here, we want to leverage this principle of decentralization and show that working with local control modules that only have access to local information is sufficient for learning stable walking behavior. The controller for the six legged robot should be learned using Deep Reinforcement Learning.

3.3.1 Reinforcement Learning

Reinforcement learning aims to solve optimization problems where an agent (here a simulation of the six-legged PhantomX robot, Fig. 3.4) takes actions in an environment in order to maximize an external reward. The goal is to find a policy $\pi(S)$ for the Markov Decision Process that returns a probability distribution over the possible actions in such a way that the expected long-term return is maximized. This is realized through interacting with the environment and directly learning from this interaction an estimator, for example, for this policy $\pi(S)$.

In the case of motor control, we are dealing with a given continuous state space which is spanned by the sensory signals and with a continuous action space which corresponds to direct motor signals. Therefore, learning operates over high dimensional continuous spaces which makes it impossible to learn an exhaustive list matching all possible states to optimal actions. Instead, in Deep Reinforcement Learning deep neural networks are used as non-linear function approximators for the policy (Arulkumaran *et al.*, 2017). The deep reinforcement learning algorithm used in our approach was Proximal Policy Optimization (PPO) (Schulman *et al.*, 2017) as it has shown to work well on a variety of problems without the need of intensive hyperparameter tuning. We used the baseline implementation of PPO provided by OpenAI Gym (Dhariwal *et al.*, 2017). It is connected using ROS2Learn (Nuin *et al.*, 2019) which enables communication between OpenAI Gym and ROS 2 enabled robots.

As a reward function, we simply used the traveled distance over a fixed time span (an episode) in order to avoid any bias due to reward shaping (Heess *et al.*, 2017). During training, the agent uses the experienced reward R_{t+1} it received

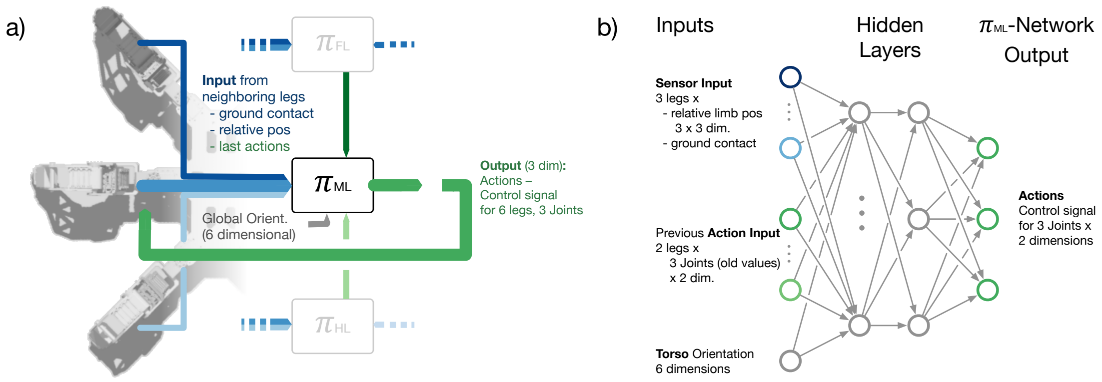


Figure 3.5: Overview of the decentralized control architecture, shown for the middle left leg (ML): a) Schematic showing local inputs to the controller. These are sensory information (blue) from the leg and its' two neighbors and last actions of the neighboring controllers to provide context. Control output (joint activations for only the respective leg) are shown in green. For each leg there is an individual controller in the decentralized structure. b) View of decentralized policy network: a neural network with two hidden layers (64 hidden units in each hidden layer). Sensory input is given as information from leg. Output is mapped to joint actions (for the decentralized approach, six such networks are used which each produce joint activations for a single leg and only receives partial sensory inputs).

from executing an action A_t in state S_t in order to optimize its' policy π. The neural network that acts as an approximator for the optimal policy is trained on these experiences. Afterwards, it should generalize and be able to infer which action leads to the highest external reward for a novel state. Updating the neural networks' weights is called an epoch.

3.3.2 Decentralized Motor Control Architecture

The decentralized architecture (shown in Fig. 3.5 a)) actually consists of six individual controllers, one for each leg. This local structure is biologically inspired from the organization of motor control systems as found in insects (Schilling *et al.*, 2013*a*). The approach is comparable to multi-agent reinforcement learning as each of these leg controllers individually controls the actions of this particular leg. As inputs (which define the controllers' observation space) each controller receives all the information of that particular leg (positions of the three leg segments, ground contact). But in addition, each controller gets local information from the two neighboring legs which differs from a multi-agent reinforcement learning approach in which all information would be decentralized and information from other agents would not be available (note, that in some multi-agent approaches this restriction is weakened and only during training all information might be accessed (Lowe *et al.*, 2017; Foerster *et al.*, 2018)). In the decentralized approach this information is structured following the body topology of the robot and taking inspiration from what we know about how information between controllers is shared in motor control in insects (currently we are simply taking a broad approach offering all local information to the learning procedure).

Furthermore, information on the orientation of the body is provided as a six-dimensional input following (Zhou *et al.*, 2019) and the last joint actions from the two neighboring legs. As a consequence, each of the six decentralized controller receives a 42-dimensional input to its policy control network.

3.3.3 A Centralized Approach as a Baseline

This approach is compared to a standard DRL single-agent approach in which a central holistic motor control structure is provided with all the information which has to come up with actions and motor control for all 18 joints of all six legs simultaneously. The input space in this case is 84 dimensional. Therefore, in the case of the centralized approach, it is guaranteed that all sensed and known information is available to the control system to make reasonable decisions. But as a disadvantage the input space is much higher dimensional which makes it more difficult for reinforcement learning to uncover on which information decisions should be based and how these should influence action.

3.3.4 Policy Network

The policy networks were setup using a similar architecture for both conditions: a policy network consisted of two hidden layers each with 64 units and using *tanh* as activation functions (see Fig. 3.5). The major difference were the input and output spaces: for the centralized approach a single network was trained with 84 input dimensions and 18 output dimensions. For the decentralized architecture six individual networks (one for each leg) were trained with each 42 input dimensions and 3 output dimensions.

Details on learning the decentralized control architecture are given in the appendix (Schilling *et al.*, submitted, 2020*a*, B.2, page 196). The article further provides detailed results for application on a simulated hexapod robot that we will briefly summarize in the next section.

3.4 Results Overview

Results on DRL of a decentralized control architecture can be summarized as, first, the decentralized approach produces viable solutions for motor control even though there is only limited information available. Second, comparing multiple learning runs shows that decentralized learning leads to control structures with a significantly better performance. Third, as DRL is a form of explorative learning, the smaller observation space of the decentralized approach helps to speed up learning considerably. Last, decentralized controllers transfer well to different terrains. These results will be briefly explained in the following.

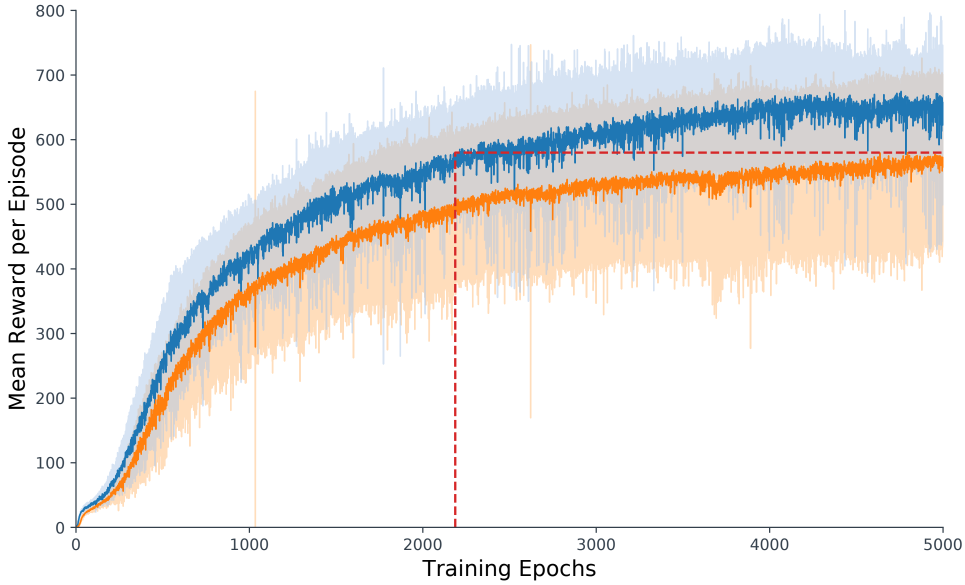


Figure 3.6: Comparing performance over training time in epochs: mean reward over all decentralized controllers is shown in blue and mean reward for the baseline centralized approach is shown in orange (shaded areas show standard deviation). Performance is measured as reward per episode (distance travelled in a single direction). Seeds were only measured up to 5000 epochs as learning appears to have converged by then. The horizontal red dashed line visualizes the maximum of the centralized approach at the end of training. The vertical red dashed line indicates when this performance level was reached by the decentralized approach (after around 2200 epochs already).

First, we consider learning to walk on flat terrain during training. As a baseline, we took a single-agent (centralized) control architecture that has access to all available sensory information and information on the last actions. Both, the baseline and decentralized approach, were trained 15 times using different random seeds. Performance over learning is shown in Fig. 3.6. Both learning approaches learned good walking behavior with quite a high velocity that looked well coordinated.

The trained controllers from the last epoch were afterwards evaluated. For each controller for over 100 individual episodes the performance (mean velocity over the episode) was evaluated. For the centralized baseline approach the mean reward was 546.70 (standard deviation 131.23). The proposed decentralized architecture consisting of six local control modules reached a performance of 657.11 (std.dev. 68.07). We performed a two-tailed Welch t-test (following Colas *et al.* (2019) with the null hypothesis that the two approaches perform equally well). The null hypothesis could be rejected (p-value of .011) and we could conclude that the decentralized architecture performs significantly better (relative effect size was large at 1.02).

The learned architectures were trained for 5000 epochs as initial tests showed that at this point controller performance was already converged for both cases. Learning progressed in several stages from simple crawling towards walking lifted

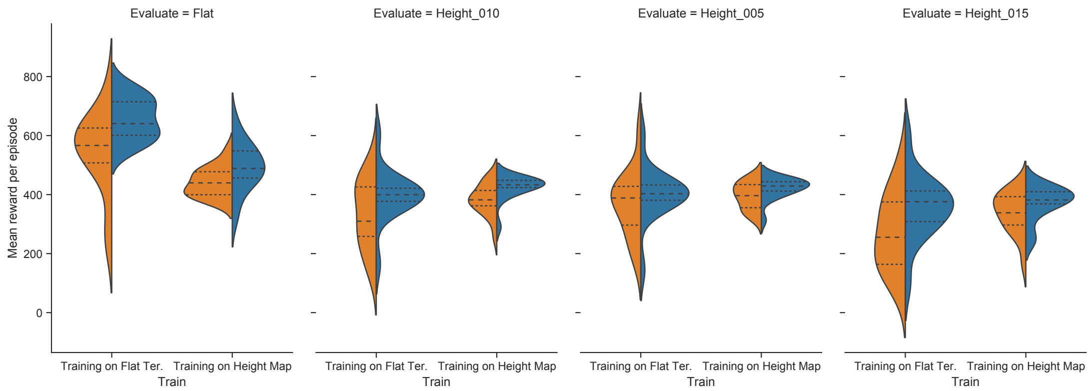


Figure 3.7: Comparison of the (approximated) performance distribution for the controllers (taken from 15 seeds per condition). Shown are violin plots for four evaluation conditions, from left to right: evaluation of walking on flat terrain; evaluation on three different height maps of maximum height of 0.05 m, 0.10 m, and 0.15 m. Inside the four evaluations, it is further distinguished between controllers trained on the flat terrain (left side) and controllers trained on the height map (right side). The violin plots show approximated distributions (and the quartiles are given inside): orange shows the centralized baseline approach and blue shows the decentralized architecture. Performance for the different seeds for each condition was measured as mean reward per episode (indicates mean velocity over simulation time).

from the ground. Comparing the two different architectures, the decentralized architecture reached a higher reward level earlier and appears to learn much faster. Compared to the mean reward of the centralized approach at the end of training, the decentralized approach reached such a level already after around 2200 epochs.

Last, we evaluated performance on uneven terrain. For this evaluation, controllers at the end of training were used. These controllers were evaluated on different types of uneven terrain for 100 episodes (using height-maps of height 0.05 m, 0.10 m and 0.15 m m that were generated using the diamond-square algorithm Miller (1986)). Pooled data for two conditions is given in table 3.1 and Fig. 3.7.

First, reward went down for uneven terrain compared to walking on flat terrain which is to be expected. For comparison, we trained controllers directly on uneven terrain (table 3.1) which shows a similar performance. Therefore, we can conclude that the controllers transfer well to novel terrains. The decentralized approach still appeared to perform better, but difference was not significant ($p = .10$).

3.5 Conclusions

The long term goal of our approach is to show how decentralized control as found in animals contributes to adaptivity of locomotion. In this chapter, we discussed results showing that Deep Reinforcement Learning of a decentralized control architecture for a simulated hexapod robot produced stable and robust walking behavior. Importantly, for the decentralized architecture mean performance

Table 3.1: Comparison of rewards between the different control architectures during evaluation. Data was collected during evaluation for 100 episodes for each controller. Given are mean rewards (and standard deviation in brackets) for each group of controllers (each group consisted of fifteen individually trained controllers). There were two different controller architectures, decentralized and centralized approaches, and two different training conditions, trained on flat terrain or on uneven terrain (using a height map with maximum height of 0.10 m). A Welch t-test was performed to compare the two controller conditions (decentralized and centralized controller) which showed significant differences between the decentralized architecture and the centralized approach in two of the three cases: For walking on flat terrain with a p-value of .011. For both approaches trained and tested on uneven terrain (expert policies, last line in the table), again there was a significant difference in favor of the decentralized approach (p-value of .023). For generalization of controllers that were trained on flat terrain and evaluated on uneven terrain, we only found a trend (p-value of .10) which might be explained through the much higher standard deviation.

Condition	Decentralized Arch.		Centralized Arch.	
	Reward	Std.Dev.	Reward	Std.Dev.
Evaluation on flat terrain	657.11	68.07	546.70	131.23
Generalization to uneven terrain	397.91	83.87	334.06	112.94
Expert policies that were trained and tested on uneven terrain	*424.30*	*40.09*	*382.42*	*51.23*

was significantly better and learning was much faster compared to a common centralized baseline approach.

As discussed in the previous chapter, decentralization and hierarchical organization appear as two fundamental characteristics of adaptive behavior. In a hierarchical organization, actions can be decomposed into sub-actions on different levels of abstraction. Decentralization in particular emphasizes that the control structure encompasses the whole nervous system and control is distributed to local control modules. There is already quite some work showing the benefits of hierarchical organization in the area of DRL and this has been argued to improve walking controllers (Hwangbo *et al.*, 2019; Frans *et al.*, 2018). In contrast, our approach for a decentralized architecture in a single agent is a first application of decentralization. It appears beneficial to combine both characteristics. We have introduced the concept of such a hierarchical and decentralized architecture (see article in the appendix (Schilling & Melnik, 2018, B.3, page 206)) for a reinforcement learning setting. From our point of view, this has the promise to show broad adaptivity as it further introduces the ability of switching between different types of behavior. This would allow a system to switch between different environmental settings and shows a type of transfer which currently already produces a viable solution in hierarchical approaches. In addition, the decentralized structure would (hopefully) lead to higher robustness and adaptivity on a very fast time scale. This requires further research and analysis of how a decentralized approach can cope with unpredictable and more diverse environments. Furthermore, application on a real robot should further test how well a decentralized architecture can deal with noise and changing body properties as well as environments.

3.6 List of Publications

This chapter gave a brief introduction and summary of three articles that were published at relevant conferences. These articles can be found in the appendix.

3.6.1 Contributions to the Thesis

- Schilling, M., Ritter, H., and Ohl, F.W. (2019), "From Crystallized Adaptivity to Fluid Adaptivity in Deep Reinforcement Learning — Insights from Biological Systems on Adaptive Flexibility". In 2019 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Bari (I).

  Appendix B.1, page 188: Discusses in detail learning on different timescales as a form of fluid adaptivity that allows to continuously adapt to changing environments and contrasts this with approaches that learn how to switch between different types of behaviors. Introduces deceptive problems in more detail. This was published at IEEE SMC (ranking: CORE B).
  Author Contributions: Conceptualization and writing original draft – MS. Investigation, writing review and editing – MS, FWO, HR.

- Schilling, M., Konen, K., Ohl, F.W., and Korthals, T. (submitted to IROS conference), "Decentralized Deep Reinforcement Learning for a Distributed and Adaptive Locomotion Controller of a Hexapod Robot".

  Appendix B.2, page 196: Application of decentralized controllers in a DRL setting. Showing detailed results that such a decentralized architecture learns robust controllers that even perform better compared to a baseline approach.
  Author Contributions: Conceptualization, writing original draft and data curation – MS. Methodology, formal analysis – KK, TK, MS. Investigation and software – KK. Writing review and editing – MS, KK, FWO, TK.

- Schilling, M. and Melnik, A. (2018), "An Approach to Hierarchical Deep Reinforcement Learning for a Decentralized Walking Control Architecture". In: Samsonovich A. (eds.) Biologically Inspired Cognitive Architectures 2018. Advances in Intelligent Systems and Computing, vol 848. Springer, Cham.

  Appendix B.3, page 206: Introduces how to extend the decentralized architecture towards a hierarchical approach in a DRL setting (implementation is future work).
  Author Contributions: Conceptualization and writing original draft – MS. Methodology, investigation, writing review and editing – MS, AM.

3.6.2 Further Related Publications

The summarized characteristics and findings have lead to two further publication:

- Kidziński, Ł., Mohanty, S.P., Ong, C., Huang, Z., Zhou, S., Pechenko, A., Stelmaszczyk, A., Jarosik, P., Pavlov, M., Kolesnikov, S., Plis, S., Chen, Z., Zhang, Z., Chen, J., Shi, J., Zheng, Z., Yuan, C., Lin, Z., Michalewski, H., Mio, P., Osiski, B., Melnik, A., Schilling, M., Ritter, H., Carroll, S., Hicks, J., Levine, S., Salath, M., Delp, S. (2018). "Learning to run challenge solutions: Adapting reinforcement learning methods for neuromusculoskeletal environments." In: S. Escalera, M. Weimer (eds.) The NIPS 2017 Competition: Building Intelligent Systems. Springer, Springer (2018), p. 121–154.

- Schilling, M., Konen, K., and Korthals, T. (accepted at BioRob 2020 conference), "Modular Deep Reinforcement Learning for Emergent Locomotion on a Six-Legged Robot".

A model is valuable not because it veridically captures some ground truth, but because it can be efficiently leveraged to support adaptive behavior.

— Botvinick *et al.* (2017, p. 27)

4 Hierarchical Internal Body Models — Grounding in Adaptive Behavior

Contents

4.1 From Adaptive to Cognitive Behavior 45
4.2 Internal Body Models 46
4.3 Mean of Multiple Computation Principle 48
4.4 Extension towards a Hierarchical Body Model 48
4.4.1 Application for Hexapod Locomotion 50
4.4.2 Application in Bimanual Reaching Task 52
4.5 Conclusions 52
4.6 List of Publications 53
4.6.1 Contributions to the Thesis 53
4.6.2 Further Related Publications 54

4.1 From Adaptive to Cognitive Behavior

In the previous two chapters, the focus was on adaptive control of behavior. Two key characteristics that were introduced were hierarchical organization and decentralization which allow for fast motor control. First, we addressed these aspects from a detailed biological perspective in insects that includes behavioral findings in restricted—but still broad—contexts and neuroscientific research. These studies shed light on the modular organization of the underlying motor system and showed that these properties allow to adapt to disturbances. Importantly, dealing with uncertainty often requires a quick response which doesn't allow for planning or even neural transmission to higher levels might be already too slow. Here, local mechanisms show an advantage as they can act on a very fast timescale. In a second step, we showed how to scale these organizational principles up using a reinforcement learning approach in order to learn novel behaviors and entrain these.

As pointed out in the introduction, we are following a dual process approach (Kahneman, 2011; Evans & Stanovich, 2013), i.e., there is one type of fast, habit-like processes that can automatically deal with control and execution of behavior (type 1). Secondly, there is another type of processes that deals with cognitive control of behavior which is assumed to be slower but takes higher level processes into account (type 2). Recently, this approach has been extended towards a mechanism level in the area of reinforcement learning in the distinction between model-free mechanisms—an adaptation to a specific niche—and model-based learning (Niv, 2019; Neftci & Averbeck, 2019). Until now, our focus was on model-free mechanisms that allow for fast actions (see previous chapters). Now, we are turning to model-based mechanisms, acting on longer timescales. The general underlying idea is to exploit a model of the external environment in planning as a form of mental simulation. In reinforcement learning, both, model-free and model-based approaches, are forms of trial-and-error learning: In model-free learning, control is shaped over time to evolve adaptive, fast reflex-based behaviors. In contrast, model-based learning offers a flexible way to mentally enact different possibilities. This has the advantage that a system can test a behavior out of its original context and apply a behavior when facing a novel situation. In this way, such model-based control complements adaptive behavior. Following our definition of cognition in the introduction, this constitutes cognitive selection of behavior. Importantly, this approach requires a form of adaptivity on the lower level to allow application of behaviors out of their original context.

Here, we are not starting from a learning-based approach. Instead, following a bottom-up approach our focus is on already existing early internal models. These are assumed to be grounded in an embodied system (Barsalou, 2008) and—following Steels (2003)—have co-evolved in service for action. One early example is an internal model of the own body (Cruse, 1999). In this chapter, we will briefly summarize the function of internal models in motor control laying the foundation for a grounded internal body model. Second, briefly introduce our recurrent neural network approach for such a body model and extend it towards a hierarchical model. Last, we will briefly summarize results showing how such a hierarchical internal body model serves reactive behaviors.

4.2 Internal Body Models

An internal model of the own body can be assumed a fundamental and evolutionary-early representation that can be found in many animals (McNamee & Wolpert, 2019). Such functional models are, on the one hand, required in motor control, for example, solving the inverse kinematic or dynamic task in goal-directed movements or a forward task in ballistic movements. On the other hand, such models are recruited in cognitive tasks as are planning ahead or observation of actions of a conspecific.

Usually, one distinguishes different types or functions of such internal models in the context of action. These different types of internal models can be found throughout the animal kingdom, even in insects (Menzel *et al.*, 2007; Dürr & Schilling, 2018). First, inverse models deal with transformations from a Cartesian three-dimensional space—as, for example, visually perceived—into an actuator or action space like joint values or muscle activations (Wolpert & Kawato, 1998). A typical example can be found in grasping and reaching movements in humans and animals (Shadmehr & Wise, 2005). This includes mechanisms in insects, for example, targeted leg movements in stick insects that are used to mediate information of footholds between neighboring legs (Cruse, 1979; Dürr & Schilling, 2018) and other goal directed movements (Page *et al.*, 2008; Honegger, 1981). For complex kinematic structures such inverse problems become quite hard or intractable. In particular, for redundant manipulators there are more degrees of freedom than required and as a consequence there are often multiple possible solutions (Bernstein, 1967). Second, forward models predict a position in space when current joint values or muscle activations are given. Such models are, for example, used in fast and ballistic reaching movements for which sensory feedback is too slow to modulate motor control and instead the predicted outcome of the movement is utilized (Wolpert & Flanagan, 2001). Again, such models can be found already in insects that use dynamic models for prediction (for example, expectation on a moving stimulus (Strauss & Pichler, 1998) or for capturing prey in dragon flies (Mischiati *et al.*, 2015); for review see (Webb, 2004)). A third type of models exploits redundancy of sensory systems which requires internal modeling capabilities for sensory integration (Wolpert *et al.*, 1995) (for examples in insects see (Wessnitzer & Webb, 2006)).

In motor control it is usually distinguished between the inverse and forward functions. This leads to very specific internal models serving one specific behavior and one specific function. A good example is the influential MOSAIC approach (Wolpert & Kawato, 1998; Haruno *et al.*, 2003), in which each behavior consists of a pairing of an inverse and a forward model. Such an approach has some drawbacks as it is not very efficient and as there are duplicate representations. Adaptation of such models seems quite problematic—when the body changes it has to alter the connected individual models for each behavior. Therefore, there is the longstanding notion of a single internal body representation (Acosta-Calderon & Hu, 2005; Hoffmann *et al.*, 2010; Arcaro *et al.*, 2019) that might subserve these functions in the context of different behaviors and tasks. There is a large body of literature on how such a body schema might be neuronally encoded in humans and animals. In general, it is assumed that configurations of body parts are encoded in a distributed and somatotopic fashion in distinct areas in the brain and that there are redundant representations (Andersen & Mountcastle, 1983; Georgopoulos *et al.*, 1988).

4.3 Mean of Multiple Computation Principle

One example for a functional internal body model is realized by our Mean of Multiple Computations (MMC) network Schilling (2011*b*). An MMC network is a recurrent neural network which can be used as an inverse model, forward model or for sensor integration. The general structure of the neural network is not learned, but setup following the MMC principle. The model encodes geometric and kinematic constraints in a recurrent neural network. This shapes an attractor space that represents valid configurations of the encoded manipulator structure. Importantly, diverse tasks can be given to the network as an input which initially act as a disturbance on a previous state of the network. As this disturbance spreads through the network, the encoded constraints will force all the activations to settle in a new valid attractor state. For example, an inverse kinematic problem can be solved by injecting a new position for the end effector into the network which requires the individual segments of the controlled structure to adapt towards new positions.

The Mean of Multiple Computations principle consists of two key ideas: first, the overall complexity of the controlled structure is broken down. As an example, we will use a three-segmented robotic arm (Fig. 4.3). When the whole kinematic chain is expressed as one single equation, the problem becomes quite difficult for traditional control approaches. In the inverse kinematic task there are multiple possible configurations in order to reach a certain target position. Instead, the MMC principle breaks this down into multiple computations that consist of local relationships between variables. While the individual equations become trivial to solve (they only consist of three variables), we end up with multiple of such computations. As the second key idea, the MMC principle exploits this redundancy. As each variable appears in multiple computations, it depends on multiple of these equations. The MMC network works in an iterative fashion: an update for a variable is calculated using all the equations that affect this variable. The different multiple computations are integrated towards a new value—this is realized as a simple weighted mean calculation.

The MMC principle has been already introduced in (Cruse *et al.*, 1998; Schilling, 2011*b*). In the following, we will explain how such a model can be scaled up towards complex body configurations as given in the case of a hexapod walker and how to apply it when dealing with revolute joint representations. A detailed introduction can be found in the article in the appendix (Schilling & Cruse, 2012, C.2, page 228, see in the article Material and Methods and Appendix).

4.4 Extension towards a Hierarchical Body Model

An MMC model is constituted of simple local relationships. As an advantage, such a model is easy to setup and derive. But with an increasing complexity of the body structure, the number of possible local relationships increases exponentially. As it doesn't appear reasonable to consider all possible relations between all kinematic

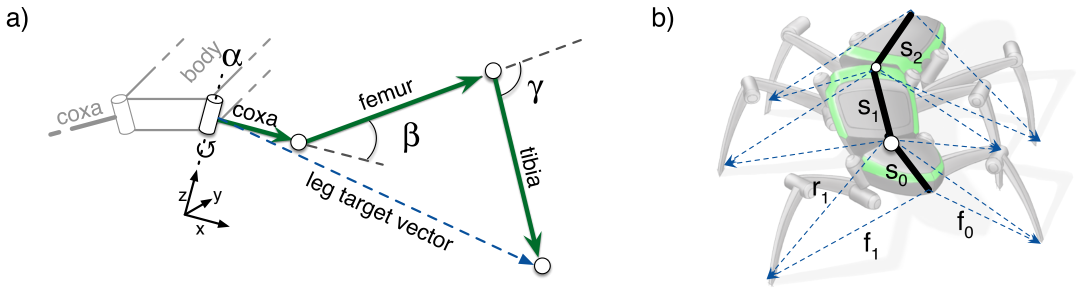


Figure 4.1: The body model for the six-legged walker is divided into two different levels. The lower layer contains six networks, each representing one leg (a). The upper layer (b) represents the body and the six legs, which are only represented by vectors pointing towards the tip of each leg. On this level, the leg is described with reference to the respective body segment. Both layers are connected via the shared representation of the target position of the leg and are implemented as recurrent neural networks. α, β, and γ denote the three joint angles.

variables, we used a hierarchical organization as a solution which introduces a form of abstraction between levels and utilizes detailed local information (Fig. 4.1).

As an example, we will consider application of such a hierarchical internal body model in the control of six-legged walking (Schilling *et al.*, 2012). For the hexapod robot, all legs that are in stance mode are coupled through the substrate. This requires a form of spatial coordination during walking. While in straight walking this requires only very little central information that are shared between all leg controllers (Schmitz *et al.*, 2008), in curve walking in stick insects legs contribute differently (see Fig. 2.7): in a narrow turn, the inner (with respect to the curve) hind leg is barely moving at all and, in contrast, the outer front leg aims in the direction of the curve, pulling the body sideways (Dürr & Ebeling, 2005). Experiments in simulation and on robots point out that this requires some form of explicit coordination and central information.

We employed a hierarchical body model that allows to distribute the computational task on two levels (Fig. 4.1 and for the network Fig. 4.2). The lower level is the leg level which comprises the detailed kinematics of a single leg (shown in green in Fig. 4.2). The higher level is the body level (blue), which comprises the description of the main body segments, e.g., the three thorax segments of an insect, and its relation to the subordinate instantiations of multiple legs. At the body level, there is no detailed information about joints. Instead, the leg is represented as a three-dimensional vector that captures the leg's contribution to support the body (this is shown by vectors connecting the main body segments (s_0 to s_2) to the feet of the six legs (l_0 to l_5), i.e., the ground contact locations. The two levels are connected through shared representations that are present in both levels (indicated by the white arrow in Fig. 4.2). Essentially, this leg target vector 'summarizes' the kinematics of the entire leg while, at the body level, it may be regarded as the desired relation between the body and substrate.

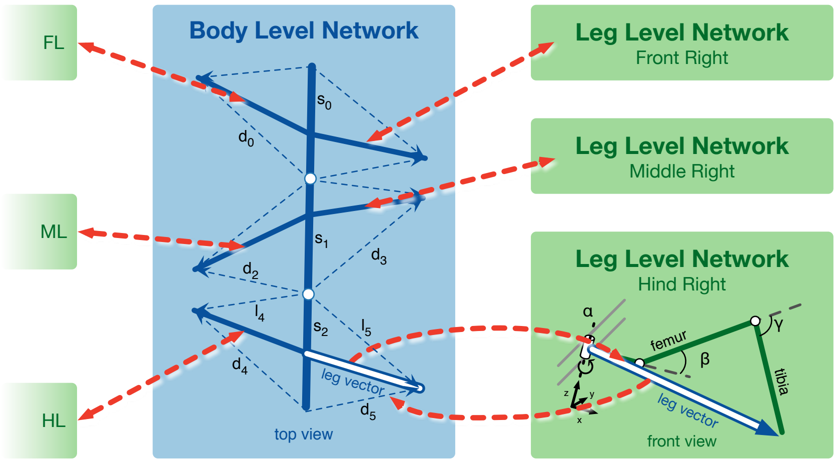


Figure 4.2: Hierarchical internal body model. The body model comprises a body level network (blue) and six subordinate leg level networks (green). Left: the body level network controls the six leg vectors, l_0 to l_5, and the three main body segment vectors, s_0 to s_2. Right: each one of the six leg level networks controls a leg with three joints and segments. The two levels are inter-connected via the shared representation of the leg vector (white arrow)

Concerning the MMC model at the lower level, each leg is described by a set of three joints and three segments. The first joint is attached to the body, while the other two connect the two distal leg segments (the trochantero-femur and tibia, see also Fig. 4.1). As a result, each leg is described by a kinematic chain with a single degree of freedom per joint on the leg level. At the body level, all legs with ground contact and all body segments are represented as three-dimensional vectors. The body model is used differently in the control of swing and stance. As a consequence, only the legs that potentially contribute to propulsion, balance and steering through body-substrate interactions are considered at this upper level of the body model. With regard to the legs in swing, all corresponding equations within the MMC network are disregarded, as if being inhibited.

4.4.1 Application for Hexapod Locomotion

The stance movement is induced into the model in the form of a passive movement, as if the body was pulled into a given direction. Computationally, this is done by displacing the front segment of the body into the direction of the intended movement. This disturbance of the body model network affects all variables contained in the equations for the connected segments. As a result, these variables are adjusted in a way which complements the enforced movement. Hence, the corresponding adjustments include three to six leg vectors (those that are currently contributing to stance). As these leg vectors are shared by the body level and the leg level networks, the induced changes 'spread' down into the leg level networks to adjust the variables of individual legs (Fig. 4.2). Thus, all joint angles are adjusted in a cooperative way, supporting the overall body movement. The procedure of making these adjustments

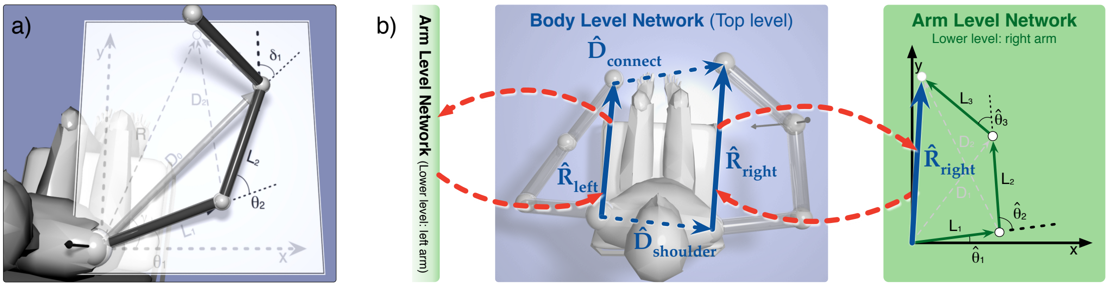


Figure 4.3: Illustration of the hierarchical body model. Shown is in a) an illustration of the lower level of the body model: it represents the structure of a single arm. The manipulator is shown which consists of three segments (L_1 , L_2 , and L_3) that are connected by three joints. The end-effector position is described by the vector R. D_1 and D_2 describe diagonal vectors. In b) the hierarchical model is shown: in the middle, the upper body level is shown. It mainly consists of the two arm vectors (shown as solid blue arrows) and the connecting diagonals. This level mediates between different higher level kinematic relations. On the right (and left) the lower detailed level of arm control is shown. Shared between those levels are the two end effector positions (solid blue arrows).

lasts for multiple iterations, as the network converges into a stable state and the resulting leg and segment vectors can be applied to control the actuators.

This internal hierarchical body model has been successfully implemented in Hector. First, in dynamic simulations the body model was used in straight and curve walking (Schilling *et al.*, 2012). It allowed Hector to navigate quite narrow curves. The stance movement of the legs was controlled through the body model. When the model was pulled at the front and forced into tight curves, the higher level came up with the complementing leg target vectors required for such tight curves. In this set of simulations the planned inter-segment drives of Hector were already realized and it showed that the additional active control of these drives provides an important contribution to negotiate tight curves. The lower level leg networks have shown to provide a robust and stable solution for calculation of the inverse kinematic function for the given target leg posture. Furthermore, the internal model has been used on the robot and has been extended for situations in uneven and rough terrain which requires an additional mechanism to deal with the control of the distance between body and the ground (Paskarbeit *et al.*, 2015).

Details on the setup of the hierarchical internal body model for the case of a six-legged robot and how this model coordinates the joint movements during locomotion are given in the article in the appendix (Schilling *et al.*, 2012, C.1, page 220). The article further explains processing between the different levels of the hierarchies and the solution of the inverse kinematic task as well as provides results for application in simulation. Furthermore, predictive capabilities of this model and coordination in a targeted reaching task are shown in (Schilling & Cruse, 2012, C.2, page 228).

4.4.2 Application in Bimanual Reaching Task

As a second example, a hierarchical internal body model has been applied in a bimanual reaching task for a human like body structure. In the case of the insect model, the leg level can be described by three joints each only having one degree of freedom and all working in a single plane, therefore describing a planar manipulator. The structure of a human arm as a manipulator is much more complex as there are overall seven degrees of freedom in three joints. This requires a different type of representation to describe the configuration of such a manipulator. Still, the MMC approach can be applied using in this case a dual quaternion representation of homogenous transformations instead of the vector description of triangular relationships (Schilling, 2011*b*; Schilling *et al.*, 2012). Such a detailed model for a human arm has been integrated into a hierarchical model following the same ideas as described above: using a detailed representation on the lower level that shares with the higher level only the target vector of the whole arm which summarizes the posture. These target representations constitute the higher level. Such a hierarchical body model and the lower level dual quaternion representation are introduced in the last article for this section in the appendix (Schilling, 2019*a*, C.3, page 248). This further shows results for a series of simulation for the case of a simple bimanual task that required computation of forward and inverse kinematics.

4.5 Conclusions

This chapter summarized the main ideas of the Mean of Multiple Computation principle. First, the body model is constituted by local relationships representing the kinematics. While in the previous chapters decentralization and local computation allowed for fast, reflex-like actions and reactions, for the body model local computations of kinematic equations—that are reduced to three variables—simplifies the control problem and makes it tractable. From a biological perspective, such local computations appear as an important part in the control of movements. One example can be found again in the locomotion of stick insects: when climbing through a twig an insect faces the difficult task of finding footholds for the legs. As a solution, information about found footholds is shared between legs starting from the front. Posterior legs are aiming at the current position of the anterior leg which guarantees a foothold as the anterior leg is already in contact with the twig (Dürr & Schilling, 2018). The position of the anterior leg affords (Gibson, 1977) a possible foothold to the posterior leg. We analyzed the computational properties of such affordance spaces and in particular the complexity of the mappings between different legs. It showed that these mappings from one three-dimensional joint angle space of the anterior leg towards the joint space of the posterior leg can be realized by a simple neural network requiring only a very small hidden layer (Dürr & Schilling, 2018). Such local computations between legs could be considered as building blocks for constituting an internal model of the whole body.

The Mean of Multiple Computation principle is based on integrating many simple local and redundant relationships. While there has been much more research in the area of encodings of spatial information of the environment, neuroscientific findings point out that there exist redundant forms of representation—and transformations—of the body which are assumed to happen in the same brain areas (Tingley & Buzsáki, 2018). One drawback of the MMC principle is that the number of the possible relations increases exponentially. As a main contribution, this chapter introduced a hierarchical organization that distributed the complexity onto multiple levels addressing this problem. This allows to deal with complex structures as given in the case of controlling two human-like arms or a hexapod walker. Application on a human-like arm required to introduce a dual-quaternion representation that allows to express joints with many degrees of freedoms. An alternative and more biologically plausible representation can be realized through population-based representation (Georgopoulos *et al.*, 1988; Morasso *et al.*, 2015; Yoo & Kim, 2017). The MMC approach has been transferred to such a form of representation as well (Baum *et al.*, 2015). Such a population-based encoding might be exploited in the future to encode noisy or faulty information in a probabilistic way.

One further drawback of the original MMC approach concerns the dynamics of the network: initially, an MMC network produces very fast movements that afterwards subsequently slow down. As one solution, MMC networks can be extended towards including dynamic relations and equations which leads to more realistic movement profiles (Schilling, 2019*b*).

To summarize: The hierarchical MMC approach provides an internal body model that is completely realized as a recurrent neural network. It allows to address inverse and forward kinematic function as well as being extended towards sensor integration. Importantly, such an internal model can be used in different contexts. This chapter detailed how such a body model serves behavior and is grounded in motor control. In the next chapter, the predictive capabilities of the internal body model will be exploited for planning ahead (Schilling & Cruse, 2017) realized as a form of internal simulation.

4.6 List of Publications

This chapter gave a brief introduction and summary of three publications that are part of this thesis and can be found in the appendix.

4.6.1 Contributions to the Thesis

- Schilling, M., Paskarbeit, J., Schmitz, J., Schneider, A., and Cruse, H. (2012), "Grounding an Internal Body Model of a Hexapod Walker—Control of Curve Walking in a Biological Inspired Robot". In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, pages 2762-2768.

  Appendix C.1, page 220: Details on the setup and processing of the hierarchical internal body model for the case of a six-legged robot. Shown are results for the control of the stance movement in forward and curve walking in simulation.

This was published in the main track of the IROS conference, one of the three leading conferences in the field of robotics (CORE A ranking).
Author Contributions: Conceptualization, methodology, investigation, analysis – MS. Software, simulation – MS, JP. Writing – MS, JP, JS, AS, HC.

- Schilling, M., and Cruse, H. (2012), "What's next: Recruitment of a grounded predictive body model for planning a robot's actions". Frontiers in Cognition, 3(383). doi:10.3389/fpsyg.2012.00383

  Appendix C.2, page 228: Provides an introduction into the MMC principle and shows predictive capabilities of this model and coordination in a targeted reaching task for an insect-like body structure. This was published in Frontiers in Cognition (Impact Factor 2.13).
  Author Contributions: Conceptualization, methodology, software, simulation, investigation, and writing – MS. Writing, review and editing as well as funding acquisition – HC.

- Schilling, M. (2019), "Hierarchical Dual Quaternion-Based Recurrent Neural Network as a Flexible Internal Body Model". Proc. of the International Joint Conference on Neural Networks 2019, Budapest (Hungary), pp. 1–8.

  Appendix C.3, page 248: A hierarchical body model for more complex manipulators. Introduces the dual quaternion representation for joints with multiple degrees of freedom and the extension towards a hierarchical model. Shows results for a series of simulations for a simple bimanual task that required computation of forward and inverse kinematics. This was published at IJCNN (ranking: CORE A).

4.6.2 Further Related Publications

The introduced local and hierarchical models have lead to further publications.

Peer-reviewed Journal Papers:

- Dürr, V. and Schilling, M. (2018), "Transfer of Spatial Contact Information Among Limbs and the Notion of Peripersonal Space in Insects". Front. Comput. Neurosci. 12:101. doi: 10.3389/fncom.2018.00101

Reviewed Conference Proceedings:

- Schilling, M. (2019), "Setup of a Recurrent Neural Network as a Body Model for Solving Inverse and Forward Kinematics as well as Dynamics for a Redundant Manipulator". Proc. of the International Joint Conference on Neural Networks 2019, Budapest (Hungary), pp. 1–8.

- Baum, M., Meier, M., and Schilling, M. (2015), "Population based Mean of Multiple Computations Networks: A Building Block for Kinematic Models". Proc. of the International Joint Conference on Neural Networks 2015, Killarney (Ireland), pp. 1–8.

- Paskarbeit, J., Schilling, M., Schmitz, J., and Schneider, A. (2015), "Obstacle crossing of a real, compliant robot based on local evasion movements and averaging of stance heights using singular value decomposition", in IEEE International Conference on Robotics and Automation (ICRA), pp. 3140-3145. doi: 10.1109/ICRA.2015.7139631

- Schilling, M., Paskarbeit, J., Schneider, A., and Cruse, H. (2012), "Flexible internal body models for motor control: On the convergence of constrained dual quaternion mean of multiple computation networks". Proc. of the International Joint Conference on Neural Networks 2012, Brisbane (AUS), doi: 10.1109/IJCNN.2012.6252846.

Cognition is the ability to relate different unconnected pieces of information in new ways and apply the resulting knowledge in an adaptive manner.

— Limongelli *et al.* (1995, p. 18)

5 Embodied Internal Simulation as Planning Ahead

Contents

5.1 Recruitment of Grounded Internal Models 57
5.2 Structure of the Cognitive Expansion 58
5.3 Results Overview 63
5.4 Conclusions 67
5.5 List of Publications 68
5.5.1 Contributions to the Thesis 68
5.5.2 Further Related Publications 69

5.1 Recruitment of Grounded Internal Models

In understanding cognition as realizing a form of planning ahead (McFarland & Bösser, 1993), internal representation play a key role. In this view, internal models are recruited in internal simulation (Hesslow, 2002) which constitutes a central mechanism that can be employed in many high level tasks, for example, in observation and anticipation of somebody's actions. In internal simulation predictive internal models are used decoupled from the body and the environment to predict what will happen. Planning ahead can be realized in this way as an anticipation of effects of possible behaviors followed by an evaluation of the anticipated outcomes. Such an internal simulation requires predictive internal models for planning ahead.

In our bottom-up approach, we extend the behavior-based system towards a system capable of using internal simulation to plan ahead. When the system runs into a novel problematic situation, it does not get stuck or use a behavior that might have dangerous consequences. Instead, the system tries a new behavior, at first, using internal simulation in order to anticipate consequences. Only then should a behavior be selected that has shown to be non harmful. Taking this ability as one holistic planning module sounds like a big developmental step which appears implausible from an evolutionary perspective. Instead, as we have discussed and shown in the preceding chapter, internal models are already required in a behaving system and can be grounded in service for action (Glenberg, 1997). Such existing internal models only require to be flexible enough and predictive to allow for internal simulation. Therefore, it is not the existence of internal models which allows for a system to become cognitive. Instead, it is the flexible recruitment of already present internal models in internal simulation that exploit predictive capabilities of the model. The neural network model introduced in the previous chapter is such a predictive model which originally subserves coordination of walking behavior. Even though this MMC model is quite a simple model, it is still flexible enough to fulfill multiple function and it can be used as a predictor in internal simulation. Such a model is not meant as providing implementational details or being realized directly in the brain. Instead it should be understood as a lower bound on the complexity, showing that already very simple recurrent neural networks can address different functions in a single model which includes predictive capabilities.

The extension of the existing behavior-based system is explained in the following section which will give a brief overview of the added structure (for details see the publications in the appendix). While the cognitive expansion realizes higher level cognitive processes, this system is still tightly connected to the underlying behaving system and recruits the adaptive motor control system in internal simulation. The structure will be motivated considering a guiding example. The result section will provide an overview of results on how the robot can solve novel locomotion problems avoiding instabilities.

5.2 Structure of the Cognitive Expansion

As an example for a novel situation, we will consider a situation in which the robot would become unstable and topple over. Such a problem can occur during climbing through twigs when, first, searching movements disturb temporal coordination as the swing movement is considerably prolonged for the searching leg (for an example see Fig. 5.1). Secondly, as a result of a searching movement, a leg is repositioned and moved further to the front which disturbs spatial coordination. As a consequence this can lead to brief instabilities as distribution of weight for the legs changed and the coordination is disturbed. At the end, we will provide a systematic analysis of such cases that can't be handled by the adaptive system as such. Importantly, such situations should not be problematic for a climbing insect which uses its' tarsi to attach to the ground. But for the robot static instabilities are dangerous as they could lead to damage to the system.

In order to cope with such a novel situation which does not trigger activation of a particular motor memory, the model has to rely on a different functionality.

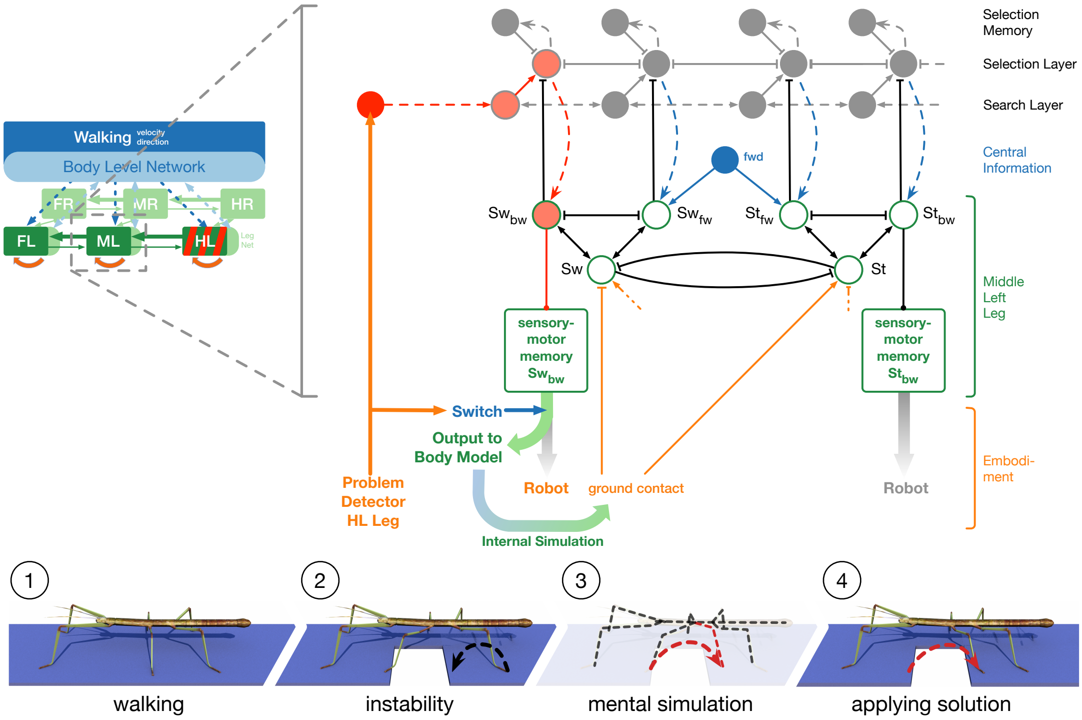


Figure 5.1: Sketch of climbing sequence which requires search for foothold and planning ahead. Lower part: System walks on narrow walkway (shown is a stylized insect figure at different points in time). Note that time is running from left to right, and robot is walking from right to left (see left most image). When a hole is encountered (image 2), a searching movement for the middle left leg is required leading to a longer step (not shown). Afterwards, lifting the hind leg causes an instability (image 2, black dashed arrow). Cognitive expansion takes over and tests different movements in simulation (third insect figure, showing an imagined movement of the middle left leg to the back). Last, this is applied on the robot (rightmost insect figure, image 4) which afterwards can continue normal walking. Upper panel shows part of the control system: Left figure visualizes detection of an instability (when trying to lift the hind left leg, HL, orange lines). Right part shows the middle left leg controller (units outlined in green), parts of the central system (blue unit, forward, 'fwd') and part of the cognitive expansion (grey units on top). First, an alternative behavior is searched for—this is induced by the problem detected in the hind left leg (shown as orange arrow), which induces activity in units marked by red shading. The three layers on top (grey units) narrow the search down to a single behavior that has not been tested before. This behavior is applied to the control system. During internal simulation motor output is routed to the predictive internal model (see lower left part, blue and green shaded arrow which shall indicate that this prediction involves local and global parts of the body model). The body model returns sensory predictions (joint and leg positions as well as problem detector signals) which allow the cognitive expansion to decide if the behavior solves the problem. If this is the case, the switch guides the output to the robot (grey arrow).

First, a predictive internal model is required. Importantly, this is already present in the current walking controller as used in the control of the stance movement as detailed in the previous chapter. Planning using such an internal model is realized as internal simulation (Hesslow, 2002). This means that instead of carrying out a novel behavior in reality, it is at first only applied on the internal body model. The body model predicts consequences of an applied behavior and these predictions can be used to decide if the intended behavior is suitable to overcome a given problem or if it should be deemed harmful. Planning ahead therefore does not rely on novel internal representation but it exploits the predictive capabilities of the internal model to infer consequences. The body itself is decoupled from the system during that time (Wilson, 2008) (see Fig. 5.1, switch). Such a form of planning ahead using an internal simulation that recruits an internal body model has been proposed by Bongard et al. (Bongard & Lipson, 2014; Bongard *et al.*, 2006). As a key difference, in our architecture searching for novel behaviors recruits the underlying motor control system. In the following, we will briefly describe how this can be realized in the neural architecture (for more details see original publications in the appendix).

As planning ahead becomes necessary whenever a novel problem arises, such problems have to be detected: A problem detector is used in two ways. On the one hand, it stops the execution of the behavior and operates the switch decoupling the body from the control process and switches towards searching for another behavior to overcome the problem. Therefore, the behavior causing the problem is stopped, too. On the other hand, the problem detector is steering the search for a solution. A search process is initiated which can be tested in internal simulation (see top part of Fig. 5.1, search and selection layer). While the current active behaviors and motivation units are excluded from the search (inhibiting connections from the green motivation units towards the corresponding units in the search layer), each problem detector is connected to at least one unit of the search layer, initiating the spreading of activation in this layer.

Finding a behavior that helps to deal with the current situation is realized in multiple stages:

- search for novel—currently not used and not associated to the current sensed context—behaviors,
- select a single alternative behavior,
- test the selected behavior in internal simulation,
- and apply the selected behavior on the real robot.

The system runs through the different stages (Fig. 5.2) each of which is associated with a motivation unit representing that processing stage (not shown here). First, the basic function of the search layer is the spreading of activation that originated from the problem detector (activate the unit in the search layer, shown as a red circle, top of the figure). When a problem has been detected, for a short period of time the activation can spread through this layer. The idea of the search layer

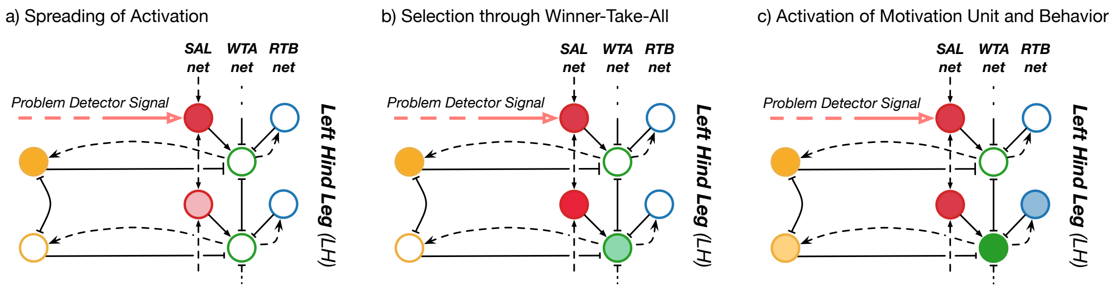


Figure 5.2: Illustration of the sequential changes of activation of units in the search layer (spreading of activation layer – SAL), selection layer (using a winner-take-all connectivity – WTA), and the selection memory layer (remember tested behaviors – RTB). When a problem occurs, the problem detector, on the one hand, stops the execution of current behavior (not shown). On the other hand, it induces search activity in the search layer (SAL, red) which indicates where the problem occurred. The activation is spreading vertically in this layer and each unit excites its corresponding unit in the selection layer (WTA, green). Importantly, currently active motivation units (yellow) inhibit these units. The selection units compete among each other producing one winning unit which in turn activates the corresponding motivation unit and behavior. The units in the selection memory layer (RTB, blue) represent which behavior has been active before and will inhibit a future activation during the selection process. Note, in this figure, color coding is used to distinguish the different layers of the cognitive expansion which deviates from the general color coding schemes of different hierarchical levels.

is that neighboring units become activated. The topology of this layer therefore determines the order in which behaviors might be selected for internal simulation. In our case, we simply assumed the structure for the walking controller as flat and organized following the body topology. After some units of the search layer became activated, a single one shall be selected in a next step in the winner-take-all selection layer. The two layers are connected in a one to one fashion. For each unit in the search layer there is a corresponding unit in the selection layer. And, as long as the spreading activation is active, the search units' activations transfer directly to the corresponding selection unit (this connection is modulated by the unit representing the search stage, modulation is not shown in the figure). In the selection (winner-take-all) network only the one unit with the highest activation stays active and inhibits all other activations after the network has converged for some time (about 10 to 20 iteration steps). When the WTA has converged, it has selected a new behavior which is close to the origin from the problem and should now be applied in internal simulation (convergence of the search and selection is shown in Fig. 5.3). The active unit of the selection layer is activating the corresponding motivation unit and, in this way, initiates the behavior. This activation of a behavior will most likely affect not only the explicitly selected behavior, but will have direct effects on the selection of other behaviors. Crucially, the behavior is not carried out on the agent itself, instead it is applied on the internal model and the model predicts the consequences of the simulation of this behavior.

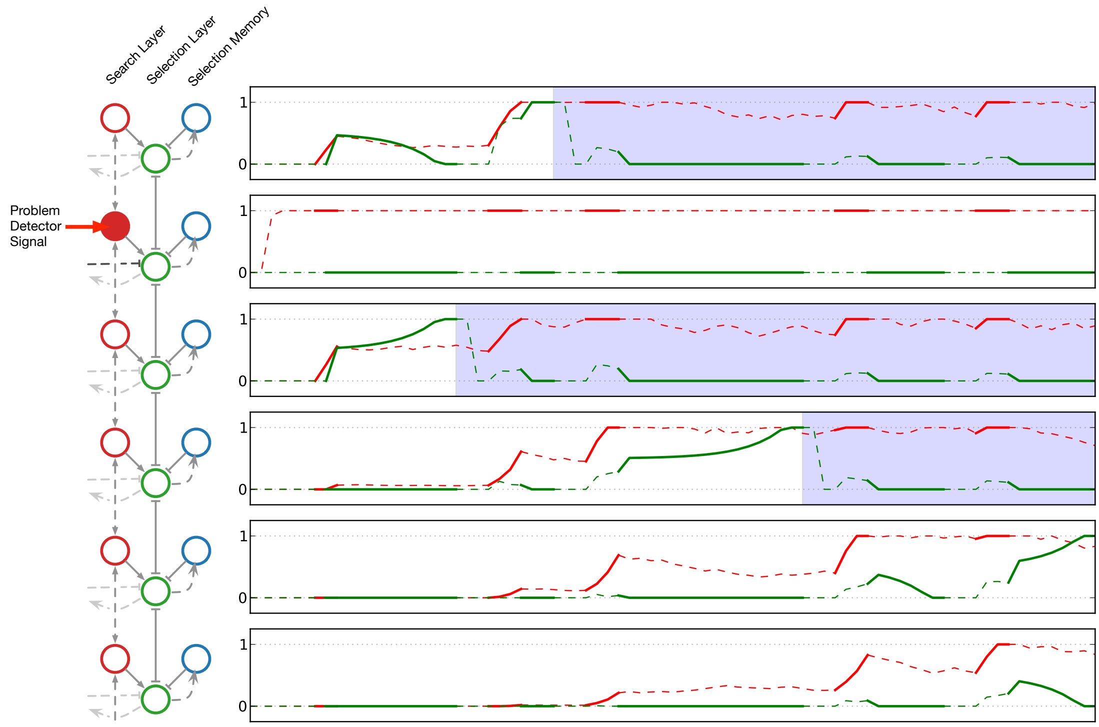


Figure 5.3: Convergence of search and selection process. Illustration of the local part of the cognitive expansion that consists of three units for each behavior (sensory-motor memory) of the control network. On the right, the corresponding temporal activations (between 0 and 1) of units of the network are shown during a run of the cognitive expansion (time from left to right). Red lines show activity of the search unit, green of the selection unit. Solid parts of the lines indicate activity during the respective stage, dashed lines show background activity of the network that does not affect the behavior of the system (time of internal simulation is not shown in figure). Blue area indicates activity of the selection memory unit after the specific behavior (shown in one row) had been selected. This inhibits the subsequent reselection (shown in green) of this action. Note, in this figure, color coding is used to distinguish the different layers of the cognitive expansion which deviates from the general color coding schemes of different hierarchical levels.

Internal simulation is only run for a prespecified time (we chose a time window that allowed for a couple of steps) and only if it is not aborted because a problem was detected in internal simulation. The internal model is equipped with problem detectors as given in the real agent. Only in this way the internal model can decide if the problem is still present or if the search has to be started again by letting the spreading activation start over. After a successful internal simulation, the selected and successfully simulated behavior is applied on the real robot in a test stage, before the robot switches back to normal walking behavior.

Detailed explanation on the cognitive extension and the overall employed architecture can be found in three publications in the appendix: Recently, we submitted an article that shows application on the real robot Hector and summarizes the work. This is provided as the first article in the appendix (Schilling *et al.*, submitted, 2020*b*, D.1, page 260) which also provides a systematic analysis on postural variations and shows how adaptive system and cognitive system complement each other. The second article (Schilling & Cruse, 2017, see D.2, page 276) introduces the cognitive extension and shows application on a simulated robot. Finally, the mentioned process stages coordinating internal simulation are explained and a detailed analysis of convergence is given in the third publication (Schilling, 2017, see article D.3, page 300).

5.3 Results Overview

The cognitive expansion has been applied in dynamic simulation and recently on the robot Hector. This section gives a brief overview of the results. On the one hand, a solution found by the cognitive expansion for the example task will be explained. On the other hand, a systematic analysis of possible postures will show how adaptive and cognitive processes are complementing each other in order to produce robust and stable behavior.

First, we are considering the exemplary situation as described before in simulation (for details see appendix (Schilling & Cruse, 2017, see D.2, page 276)). The agent is standing in an awkward posture which was induced by performing long steps far to the front for the middle left leg and hind right leg (Fig. 5.4 shows positions of the legs along the body axis over time). Its left hind leg is far to the back and has arrived at the end of the working range. Therefore, in the next moment it has to produce a swing movement and shall be lifted from the ground. But as the other hind leg and the middle left leg are moved far to the front, the hind left leg can not be lifted without the agent toppling over backwards (similar to the posture shown at the bottom of Fig. 5.1). This problem could be detected by a problem detector in different ways. A walking system could be equipped with an internal stability sensor which checks for the stability of the system at all time. Such a detector can be easily realized in the internal body model and most real robotic walking systems use similar systems for safety reasons. There are other solutions possible which are more biologically plausible, e.g., insects possess multiple load sensors and when an insect tries to lift a particular leg it could recognize that this leg is not unloaded as it has to support the body. In our case, we used a simple stability detector in the internal model as through this solution the problem detector is already present in the internal model and can be accessed as well during internal simulation. As the agent recognized in this situation that he had ran into a problem, the current behaviors were stopped, the robot was slowed down and the body itself was decoupled from the control process. In addition, the problem detector activated the corresponding swing movement unit of the hind leg in the search layer and the whole control network entered the spreading activation stage. Neighboring units became activated and through the WTA structure one was chosen. The corresponding selected behavior was then tested in internal simulation. Fig. 5.4 shows leg positions over time with internal simulation shown as red and green shaded area. While the first chosen movement (shown in red area) didn't help to re-coordinate the leg positions, for the second internal simulation a swing movement backwards for the middle left

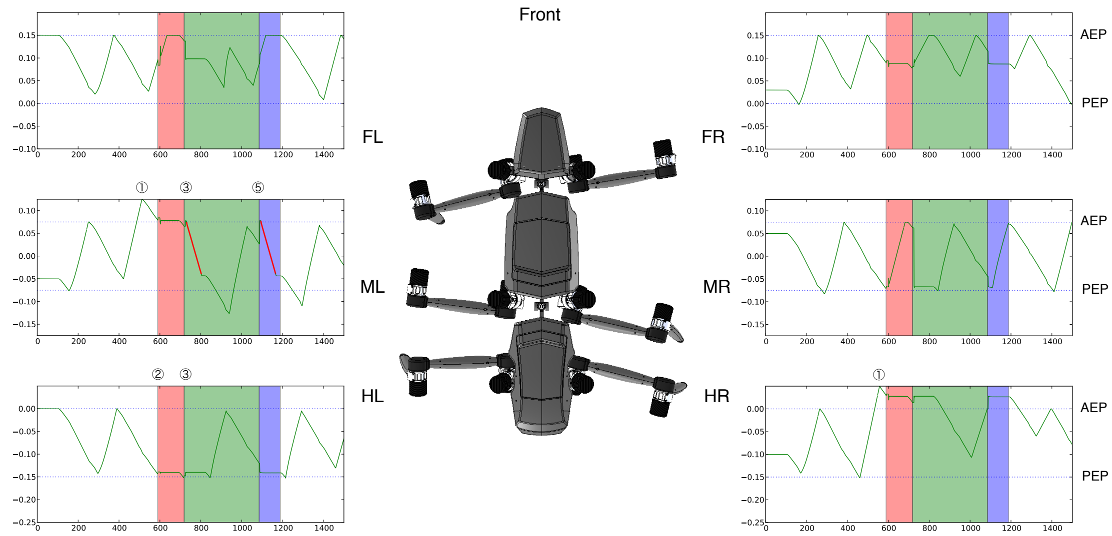


Figure 5.4: Cognitive solution to an unstable posture for a simulated robot: Position (y-axis) of the individual legs over time (x-axis). Green lines show the position of each leg over time—positive values are towards the front of the walker. Ordinate is given in cm with the origin in an intermediate leg position. The blue dashed lines indicate the average extreme positions: The Anterior Extreme Position (AEP) is the target position for the swing movement and is fixed during forward walking. The Posterior Extreme Position (PEP) indicates the position at which a leg controller initiates a swing movement on average and switches from stance to swing (note that the coordination rules act on the PEP and shift the PEP forward or rearward to organize the overall behavior which is not shown in the figure). Shortly after the left middle and right hind leg performed swing movements that pointed very far to the front of the working range (1), the walker became unstable (2) when trying to lift the left hind leg. Therefore, internal simulations was started (highlighted in green and red) during which motor commands were routed to the internal body model, the leg positions of which are shown. First (highlighted red), an unsuccessful behavior was tested: a stance movement which had initially no effect as the agent was stopped. But when the agent accelerated again (after 100 iterations) the problem was still present and the agent became unstable (3). As a second trial, a backward swing movement of the middle left leg was tested via internal simulation (green highlighted area; the swing movement in the unusual direction is plotted as a red line). Afterwards (5) the solution found was tested on the real robot (highlighted in blue) showing that walking continued successfully.

leg was selected (green shaded area, movement is highlighted as red line). This unloaded the hind left leg and allowed to continue with normal walking afterwards. It is important to note that such a selection of a behavior out of context disturbs the coordination of all legs. But this does not pose a problem for the whole system as lower and higher levels are tightly interconnected and behavior emerges out of this interaction. The lower level adapts to this disturbance of coordination and a coordinated leg pattern emerged again already during the next two steps. Last, the tested behavior was applied on the robot and solved the awkward posture.

A similar problem was applied on the real robot, but in a real behavioral context as found during climbing (details see appendix, (Schilling *et al.*, submitted, 2020*b*,

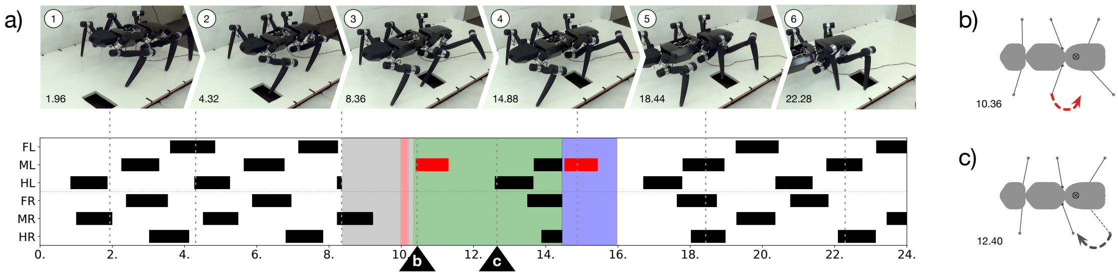


Figure 5.5: Robot climbing over a hole in a walkway, reestablishing stable walking through internal simulation. Robot Hector walks on a walkway (from right to left, vel = 0.016, fast tetrapod gait) that contains a hole (black rectangle): Upper panel shows a perspective view (images 1–6). Lower panel shows footfall pattern, swing movements are indicated as black (or red) bars for the six legs (given on y-axis) over time (x-axis, running from left to right, in seconds). While the front left leg (FL) performs a swing movement over the hole (image 2, time = 4.32 s), the second swing movement of the middle left leg (ML) was reaching into the hole followed by a searching movement of the swinging leg (at about $6 - 7$ s, not shown). As a consequence, the middle left leg was moved far to the front. When the hind left leg had to be lifted (image 3, time = 8.36 s), the robot got unstable. Therefore, the robot was stopped and started the search procedure (grey area in footfall pattern). The internal simulation (light red area) showed that this attempt was unsuccessful and was therefore aborted. In the second attempt (grey and green area), the middle left leg performed a step backwards, marked by red bar for indicating back swing in footfall pattern). This action is also illustrated in subfigure (b) (shown is a top view of the body model and the movement of the middle left leg, orange dashed arrow). It unloaded the hind leg, which could then perform a normal swing (illustrated in subfigure (c), black dashed arrow). Then (simulated) normal walking is resumed (still green area). As internal simulation turned out to be successful this action was performed out of context on the robot (blue area, image 4) and the robot continued normal walking (white area, and images 5, 6).

D.1, page 260)). In this case, the robot Hector was walking on a walkway with a hole. When a leg stepped into the hole during a swing movement, a search movement was automatically triggered and the leg searched for a foothold by moving the leg further to the front. In the specific case shown in Fig. 5.5, the middle left leg stepped into the hole which required repositioning of that leg to the front. As a consequence the robot ended up in a similar configuration as explained above. Processing progressed the same way as described before: when the system tried to lift the hind left leg, the robot became unstable and would have toppled over. A problem detector recognized the problem and safely stopped the walking behavior. In addition, the cognitive expansion took over, first, searching and selecting a different behavior out of context (shown in grey areas) and afterwards testing behaviors in internal simulation (red and green shaded areas). In this case, the system found the same solution as described above (in other cases, repositioning of the hind right leg was used as a solution). Last, this was applied on the real robot (shown in blue shaded area in Fig. 5.5).

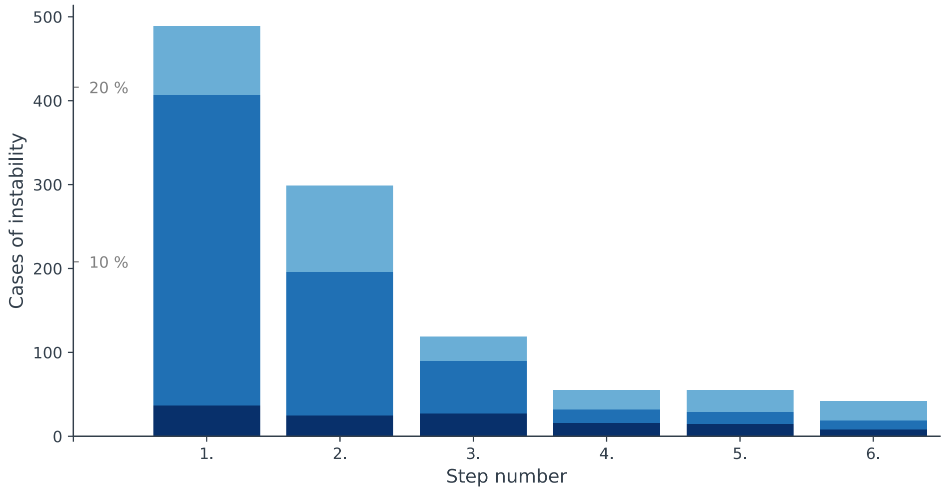


Figure 5.6: Instabilities of the control architecture (without using the cognitive extension) when forced into systematically varied starting postures. Abscissa: number of robot steps. Light Blue: duration of instabilities between 10 ms and 100 ms; blue: duration longer than 100 ms, but only appearing during a single step cycle of that particular leg; dark blue: long instabilities (longer than 100 ms) and found in subsequent steps of the walking robot. Shown for high velocity (0.020).

Further runs on the robot are described in the original article and can be summarized: for small disturbances the system was not affected and stable walking patterns emerged. For larger disturbances the robot became unstable and the cognitive expansion had to take over. As the problem detector induces a search close to where the problem occurred, usually, a preferred solution was found and only seldom another solution was selected. Importantly, applying an action out of context necessarily breaks the coordination pattern that has emerged up to that point in time. In most cases, this was not problematic as the system converged very quickly again towards a coordinated walking pattern that allowed for stable walking. This is due to the adaptivity of the underlying decentralized control architecture. For other cases this required further intervention later-on.

Last, we tested systematic variations of the starting posture, at first for the decentralized architecture without using the cognitive expansion to obtain a coarse estimate of how many different starting positions lead to a stable walking pattern. In a second step, we compared this to the architecture that includes the cognitive expansion and analyzed how the cognitive expansion dealt with the problematic cases. For each leg four different starting postures were assumed that were equally spaced from the front (anterior extreme position) towards the back (directly in front of the posterior extreme position). This poses a quite challenging task for a controller, as in many cases phases between leg controllers initially differed substantially from a typical, stable walking pattern. Overall, we ended up with 2080 different starting postures (4^6 minus all symmetric configurations with respect to the body axis). The simulated robot was subsequently initialized adopting the different starting postures.

Table 5.1: Overview of occurrence of instabilities resulting from 2080 different starting postures. These were counted with respect to step cycles (after initialization) and we distinguished very brief instabilities that (on inspection of simulation run) did not require any intervention and longer instabilities that required the cognitive expansion.

Step number	**1**	**2**	**3**	**4**	**5**	**6**
Duration of instability						
10 – 90 ms, light blue	82	103	29	23	26	23
Instability $\geq$ 100 ms (but single step), blue	370	171	63	16	14	11
Instability spanning mult. steps, dark blue	37	25	27	16	15	8

Afterwards, the controller were started with the defined high velocity and the robot started to walk. A posture was determined unstable when the center of gravity left the polygon spanned by the standing legs. To illustrate the difficulty of the task: in normal walking (at a fast velocity) neighboring controllers are assumed to be in anti-phase relation. In contrast, from the 2080 initial postures, 1216 are defined with in phase relations between neighboring legs (even when excluding middle leg symmetries still 928 initial postures are characterized by phase relations that would cause instabilities when maintained during walking). The number of instabilities, and correspondingly the durations of instabilities decreased strongly during the first couple of steps (Fig. 5.6 and table 5.1; data from 2080 different starting configurations). After three steps, mostly only brief static instabilities could be observed. For an intermediate walking velocity, there were less frequent instabilities, but the same trend was observed: over time the controller emerged towards stable gaits. Importantly, when applied on the complete system that includes the cognitive expansion, the cognitive expansion took over in unstable situations. It allowed to search in a safe way in internal simulation for a behavior that lead to a different posture in which load was better distributed between legs. In this way, the cognitive expansion facilitated stable walking and resolved unstable postures.

Detailed results on the robot and systematic variation of starting postures are given in the appendix (Schilling *et al.*, submitted, 2020*b*, D.1, page 260). First results in simulation are shown in detail in (Schilling & Cruse, 2017, see D.2, page 276).

5.4 Conclusions

Following a bottom-up approach, we extended our system towards a minimal cognitive system that is capable of planning ahead realized as internal simulation (Schilling & Cruse, 2017). This enabled the hexapod robot to, first, still show adaptive behavior and produce robust walking behavior across quite a spectrum of different walking contexts. Secondly, when the robot was forced into a dangerously unstable posture, the cognitive expansion was able to take over. A suitable solution was found in readjusting the posture by choosing an alternative behavior outside of its' original context which was initially only tested in a safe internal simulation. As a last step, this behavior was applied on the real robot.

While the system shows a robustness towards variations in the environment, there are still some limitations. Currently, the cognitive controller is only searching and considering a single behavior as a solution to a given problem and not chaining multiple behaviors towards a sequence. This has shown to resolve unstable postures. But for climbing in an environment with very sparse footholds, the system should be extended and should become able to handle more complex variations allowing for sequences of behavior. Scaling towards more difficult problems would also require learning of found solutions which is future work.

It is important to note that the details of the cognitive extension of the control system are not based on assumptions on the structure of an insect control system. There is only little research on planning capabilities in insects (Giurfa & Menzel, 2013; Menzel *et al.*, 2007) which addresses how predictive memories are employed for guiding goal-directed movements (Card & Dickinson, 2008) or for generalization of behaviors (Loukola *et al.*, 2017). Instead, the extension is motivated from the broad behavioral and neuroscientific support for the flexible recruitment of internal models in cognitive tasks in other animals including humans (Anderson, 2010; Chersi *et al.*, 2013; Ólafsdóttir *et al.*, 2015). In this context, Embodied Cognition identified internal simulation as a key mechanism that allows to exploit existing internal models in cognitive tasks: ranging from observation over planning ahead to use of language (Gallese & Lakoff, 2005; Pulvermüller, 2018).

5.5 List of Publications

This chapter gave a brief introduction and summary of three publications that are part of this thesis and can be found in the appendix.

5.5.1 Contributions to the Thesis

- Schilling, M., Paskarbeit, J., Ritter, H., Schneider, A., and Cruse, H. (submitted), "From Adaptive Locomotion to Predictive Action Selection—Cognitive Control for a Six-Legged Walker".

  Appendix D.1, page 260: Summary of the structure of the cognitive expansion and application on the real robot Hector in a climbing task (Schilling *et al.*, submitted, 2020*b*, D.1, page 260). In addition, provides a systematic analysis on postural variations and shows how adaptive system and cognitive system complement each other.
  Author Contributions: M.S. and H.C. designed and performed research as well as analyzed the data. M.S. wrote simulation software and developed concept of the model. Simulation experiments were performed by M.S. and experiments on robot were performed by J.P and M.S. M.S. wrote the paper and H.C., H.R., and A.S. contributed to the writing.

- Schilling, M. and Cruse, H. (2017), "ReaCog, a Minimal Cognitive Controller Based on Recruitment of Reactive Systems". Frontiers in Neurorobotics 11(3). doi: 10.3389/fnbot.2017.00003

  Appendix D.2, page 276: Introduces the cognitive expansion and demonstrates the application for the case of an awkward posture on a simulated robot (Schilling *et al.*, submitted, 2020*b*, D.2, page 276). This was published in Frontiers in Neurorobotics (Impact Factor 3.00).
  Author Contributions: MS and HC laid out the concept, designed the model and analyzed the data. MS carried out the implementation. MS and HC wrote the manuscript. Writing review and editing all authors.

- Schilling, M. (2017), "Old Actions in Novel Contexts — a Cognitive Architecture for Safe Explorative Action Selection". Proceedings of the Artificial Intelligence and Simulation of Behaviour Conference (AISB 2017), Bath (UK).

  Appendix D.3, page 300: Analyzes in detail the stages coordinating cognitive processing and shows their convergence (Schilling, 2017, see article D.3, page 300).

5.5.2 Further Related Publications

The work detailed in this chapter has lead to further publications.

Peer-reviewed Journal Papers:

- Schilling, M. (2016), "Lose a leg but not your head – a cognitive extension of a biologically-inspired walking architecture". Procedia Computer Science 88, pp. 102-106.

- Schilling, M., Rohlfing, K., and Cruse, H. (2012), "Prediction as internal simulation: Taking chances in what to do next". Frontiers in Psychology, 3(405). doi:10.3389/fpsyg.2012.00405

Book Chapters:

- Cruse, H. and Schilling, M. (2018), "Getting cognitive". In B. Bläsing, M. Puttke and T. Schack (Eds.), The Neurocognition of Dance, second edition.)

- Cruse, H., and Schilling, M. (2015), "The Bottom-up Approach: Benefits and Limits". In: Open Mind, Metzinger, T., Windt, J. (eds.); Frankfurt/M.: MIND Group Frankfurt/M.

[The] flexible, combinatorial aspects of planning will form a critical underpinning of what is perhaps the hardest challenge for AI research: to build an agent that can plan hierarchically, is truly creative, and can generate solutions to challenges that currently elude even the human mind.

— Hassabis *et al.* (2017, p. 253)

6 Discussion and Conclusion

Contents

6.1 Key Principles of Hierarchical Control 72
6.1.1 Lower Level Motor Control 73
6.1.2 Mid-level Action Control 75
6.1.3 Higher Level Control – Planning 76
6.2 Conclusion 77
6.3 List of Publications 80
6.3.1 Contributions to the Thesis 80
6.3.2 Further Related Publications 81

With the recent success of Artificial Intelligence (AI) methods in some tasks—often in quite restricted areas as are game playing or image recognition (Arulkumaran *et al.*, 2017; LeCun *et al.*, 2015; Mnih *et al.*, 2015)—there is now as well a growing interest in applying such methods in real-world settings and in artificial agents. But broadening the scope of approaches—towards more diverse application areas, introducing other agents that make the environment more unpredictable and allowing for noise on sensory as well as other levels of processing—dramatically increases the difficulty of the problem and requires qualitatively different approaches. Here, again natural systems can provide inspiration and are starting to influence current research (Hassabis *et al.*, 2017; Neftci & Averbeck, 2019; Ullman, 2019; Storrs & Kriegeskorte, 2019). In one recent perspective, Merel *et al.* (2019*a*) in particular provided a view on the hierarchical organization of motor control in mammals and how this can influence the design of artificial systems (Merel *et al.*, 2019*b*). They provide a set of characteristics which we will use in this chapter to analyze our proposed architecture. First, their key principles will be briefly introduced. Secondly, following their distinction on three different levels of processing we will discuss our architecture with respect to these three levels and their characteristics.

This will, on the one hand, highlight how our system realizes key characteristics with respect to hierarchical organization which is in agreement with their argument. As our architecture follows detailed neuroscientific findings from insects this further demonstrates that such a structure constitutes basal organizational principles and insects offer here an ideal model system as they allow for analysis on different levels of detail. Our work adds and complements the characteristics of Merel *et al.* (2019*a*), in particular with respect to decentralization and embodiment. On the other hand, as they turn towards higher levels of control, they directly provide a sketch of how to further extend our system.

6.1 Key Principles of Hierarchical Control

Merel *et al.* (2019*a*) provide a perspective on research in motor control and artificial control systems. Their goal is to point out how such a view can be mutually beneficial for both research fields. In motor neuroscience, focus on single behaviors and movements has allowed to develop and validate theories as optimal feedback control (Todorov & Jordan, 2002). Optimal feedback control has become an influential general paradigm in which a movement can be described with respect to a specific objective or cost function. Similarly, optimization became influential over the last years in the area of deep reinforcement learning. As one example, Heess *et al.* (2017) trained simple feedback controllers for locomotion of a simple simulated walking agent in an end-to-end fashion. While a single feedback controller showed a specialized behavior, the overall system became robust through a hierarchical organization in which—on a higher level—switching of controllers was realized. Merel *et al.* (2019*a*) focus in their perspective on such hierarchical organization of motor control systems in animals and artificial systems. They specify key characteristics of motor control across different levels of such a control hierarchy. As current Artificial Intelligence approaches still struggle to scale towards more complex systems, real world scenarios, and more diverse as well as noisy situations, they propose that artificial control approaches can benefit from following these core principles. We will provide a brief overview of their principles:

- **Temporal abstraction:** As there are different levels in a hierarchy, each level is operating on a different temporal scale (Sutton *et al.*, 1999). This simplifies specification of behavior as it can be composed on a higher level out of more fine-grained building blocks on the lower level.

- **Multi-joint coordination:** In animals, muscles and joints can be grouped together that move in a coordinated fashion (Flash & Hochner, 2005; Latash *et al.*, 2007). Such motor synergies introduce another form of abstraction that simplifies learning or exploration of new skills.

- **Information factorization:** Different levels of a control hierarchy are as well associated with different levels of representation. Information is factorized differently on the different levels. As higher levels use, for example, pre-processed information stemming from sensory inputs and processing, information becomes summarized and more abstract.

- **Partial autonomy:** Elements on the lower level can function autonomously without requiring the need of top-down input driving these modules. This relates to our explicit notion of decentralization and the general idea that local control mechanisms can respond quickly. Lower level modules are acting concurrently and overall control is mediated between the different levels (Brooks, 1986): in the bottom-up direction information is factorized for higher levels while, in the top-down direction, higher level control modulates lower levels which addresses details of control.

- **Modular objectives:** A crucial distinction in cognitive systems differentiates between processes and representation. This is addressed by Merel *et al.* (2019*a*) in their modular view: on the one hand, information is factorized in modular representation and, on the other hand, processes operate in a concurrent manner. One particular focus in motor neuroscience and AI approaches to motor control is on objective functions for which Merel *et al.* (2019*a*) extend their modular view. As there are different modular processes acting on specific (partial) information, these should—in their view—be considered to serve different specific objectives.

- **Amortized control:** Often repeated execution of specific movements should be realized computationally efficient, for example, through caching successful solutions for a given context.

Merel *et al.* (2019*a*) put a focus on hierarchical organization and broadly differentiate three levels. While they review work on mammals, they discuss general principles which—at least for lower and intermediate levels—apply for other animals as well (Dickinson *et al.*, 2000; Webb, 2020). These will be discussed in the following in relation to our architecture and how this relates to their proposed characteristics.

6.1.1 Lower Level Motor Control

Merel *et al.* (2019*a*) highlight the modular structure of motor control on the lower level. Spinal circuits in mammals are directly responsive to sensory feedback without higher level inputs (Bizzi *et al.*, 1991), as shown in decerebrated cats (Grillner & Zangger, 1979; Whelan, 1996). This general notion of distributed motor control is reflected in our approach (Fig. 6.1) as decentralization constitutes a main organizational principle: local control clusters on the lower level (shown in green) act on local information. This local information is directly factorized based on the availability of only nearby sensory inputs. These decentralized control modules are partially autonomous as they are driven by local signals, but can be modulated by

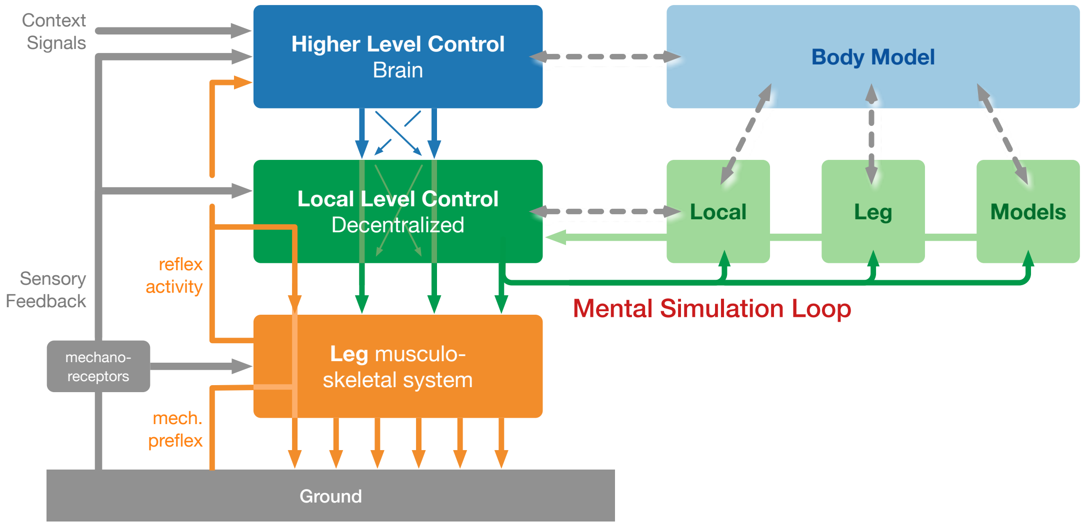


Figure 6.1: Overview of the motor control architecture: The architecture realizes adaptive and cognitive behavior. On the left, the different control levels are shown and how these interact. Importantly, there are highly parallel connections which highlights the concurrent and decentralized structure of the system. On the right, underlying internal models that serve behavior are shown in light shaded colors. Gray dashed arrows between the two columns signify that internal models are grounded and recruited in motor control. The single green arrow to the right indicates that during mental simulation the motor control information is rerouted towards these predictive internal models starting a mental simulation.

higher levels. Overall behavior emerges from the interaction of the autonomous control mechanisms. In our robotic experiments we could show how such a system produces robust behavior which further strengthens the argument that modularity directly benefits adaptivity (Clune *et al.*, 2013; Lipson *et al.*, 2002).

From a dual process perspective (Kahneman, 2011; Evans & Stanovich, 2013) the lower levels realize an automatic process that allows for fast responses. Such a solution is computationally efficient and allows without much effort execution of stereotypic movements and behaviors—it amortizes control. Learning in a dual process theory is assumed as entraining a novel behavior—that required effort and higher level processing—on the lower automatic level. Memorizing of found solutions is not addressed currently in our architecture and will be future work.

We are using the notion of decentralization throughout this thesis as it further highlights that there is not an arbitrary division in modules, but that this follows embodied constraints. As transduction times delay sensory information, local control has to be responsible for fast responses. This leads to decentralized modules that correspond to the body structure as realized in our leg controllers. Such an organization relates and integrates close-by sensory information as well as actuators forming multi-joint coordination patterns. In our architecture, we have seen this on the processing side in the distribution of control onto the leg level. But this is as well reflected in a representational perspective—relating back to information factorization—in which for the proposed internal body model representation the detailed coordination of the multiple joints of a leg is realized on the lowest level of the internal body model. The higher level of representation summarizes these information and hides details unnecessary for action selection.

Finally, Merel *et al.* (2019*a*) ask for models of animal behavior in physical realistic environments which should include more diverse contexts as well as noise when interacting with the real world. We agree with the importance of applying control approaches on real robots and in real settings. But we think it is important to explicitly acknowledge the need and possible advantage of embodiment. In an embodied perspective motor control not only includes control principles as described by Merel *et al.* (2019*a*), but extends to bodily properties and the interaction between all these. These bodily properties can simplify control problems or directly solve them. For example, elastic properties of muscles allow to instantaneously deal with small disturbances and protect joint actuators. Detailed work on insects—as discussed in this thesis and applied as control principles in our architecture—has highlighted the contribution of such embodied properties in locomotion. This further complements hierarchical control on the lower levels.

6.1.2 Mid-level Action Control

An intermediate level of action control is used by Merel *et al.* (2019*a*) in order to highlight the classical distinction between execution of movements and selection of different actions. Importantly, their disambiguation of three different levels is not meant to propose that there is such a strict order and that there are exactly three levels. Rather there are qualitatively different levels and processes that show a specific form of abstraction. One such qualitative difference is that on an intermediate level, low level motor primitives are subsumed (amortized control) that share a common objective (modular objectives). As mentioned above, these are partially autonomous and the intermediate level selection process is sensory driven. Our adaptive architecture adheres to these principles of hierarchical organization. Action selection is realized in the Motivation Unit network as a local competition of possible behaviors on an intermediate level. It is depending on the current state of the system and driven by sensory inputs, for example, in switching from swing to stance when a leg touches the ground. Merel *et al.* (2019*a*) point out that the intermediate level of action selection and the lower level of detailed execution can realize complex patterns without the need of higher levels (which would correspond to cortical areas in mammals). The Walknet approach shows such an emergence of complex adaptive behavior that can deal with quite diverse situations. It is driven and modulated by simple higher level signals as are movement direction or walking velocity. One difference compared to the view proposed by Merel *et al.* (2019*a*) is that in our approach action selection is realized on multiple different levels and explicitly implemented as a simple Winner-take-all structure (for example, there is a competition on the leg level for selecting a swing or stance movement, but in the extension towards forward and backward walking this selection is already represented as a hierarchy on the leg level and there is a further local competition

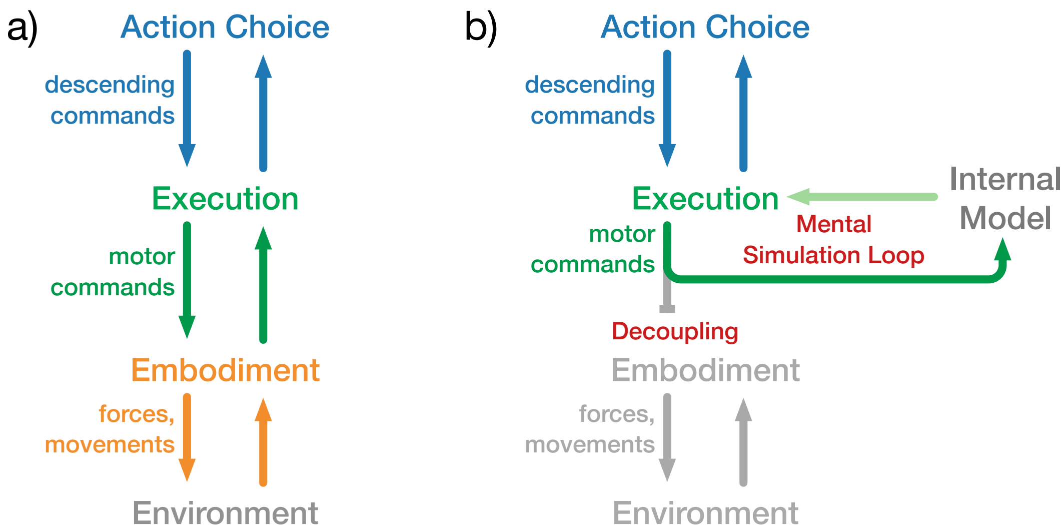


Figure 6.2: Overview of the motor control architecture: The architecture realizes adaptive and cognitive behavior. a) provides an abstract and simplified schematic that visualizes processes between different levels of hierarchies in adaptive behavior. b) shows how decoupling of the body allows to realize a form of planning ahead as mental simulation which constitutes cognitive behavior.

on a higher level between different behaviors). Furthermore, one open question for Merel *et al.* (2019*a*) is how different levels and modules interact and share (factorized) information. In the Walknet approach there is information flow realized between levels. But in addition, there is local coordination between modules on the same level: the leg controllers are not coordinated through a higher level synchronization signal, but instead coordination emerges out of simple local coordination influences.

6.1.3 Higher Level Control – Planning

Merel *et al.* (2019*a*) describe the goal of higher control levels as to provide flexibility for unrehearsed movements. They connect this to Bernstein's notion of dexterity that aims at *"finding a motor solution for any situation and in any condition"* (Bernstein, 1996, p. 21). In their view, higher level control overrides automatic responses of the lower level in such scenarios. This fits well with the distinction in dual process types (Evans & Stanovich, 2013) we have used in the introduction: automatic lower level control processes are in our architecture complemented by reflective controlled processes (see Fig. 1.2 for an overview of characteristics of the two types of processes). In both views, this introduces an effortful search for solutions to a given problem which is realized as a mental simulation in our architecture (Fig. 6.2). Merel *et al.* (2019*a*) put a particular focus on sensory input to higher level control. Their notion of information factorization that induces a form of abstraction in the bottom-up processing of sensory input can be found in how sensory input is processed in our hierarchical control architecture and is made explicit in the internal body model. Furthermore, they point out that higher level control receives a multitude of sensory inputs and, in particular, this extends to more distant information. They use visual information and how this can guide motion as an example. While in general their examples mostly focus on manipulation tasks, this extends as well towards locomotion. In biological motor control, visual guidance modulates lower levels. Currently, there is no visual input used in our system. But high level motor control in the architecture is realized as a modulation of—and projection onto—the lower levels as these are recruited in tasks. This grounding of higher levels in lower control levels is a key characteristic of our approach. It is our goal for future work to extend our model towards, on the one hand, visual—pre-processed—input that can be integrated across different levels of motor control, and, on the other hand, towards more distal information that would extend the internal model of the own body towards a representation of the environment (for example, in navigation (Hoinville & Wehner, 2018)).

6.2 Conclusion

We have devised a decentralized and hierarchical control model in this thesis which we applied for locomotion on the six-legged robot Hector. The organization of the architecture is based on key principles of motor control as found in animals: In particular, a hierarchical and decentralized structure is realized. Control is distributed onto different levels and processed concurrently which allows to respond on different time scales. On the one hand, this allows for very fast reactions. In our embodied approach this includes bodily properties that can directly absorb disturbances. In this view, motor control encompasses body, control system, and interaction between these. The lower control levels are based on detailed experimental findings of walking in stick insects. We demonstrated that this leads to adaptive behavior as the robot can quickly and automatically react to disturbances and stable behavioral patterns emerge from interaction with the environment. On the other hand, on a much longer timescale, the system is extended towards a minimal cognitive system that is able to handle novel, problematic situations. For this purpose, we have extended the system towards a cognitive system that allows as an additional mechanism to come up with changed behaviors in novel contexts. The cognitive system exploits the predictive capabilities of an internal body model that was already required in the behavior-based approach for spatial coordination during curve walking. In a form of internal simulation, existing behaviors are applied out of context in unstable, problematic situations and routed to the internal body model while being decoupled from the body. The body model that is realized as a recurrent neural network provides predictions of the behavioral outcomes. This allows the system to test potential behaviors in a safe way and only apply those that appear non-harmful. But such a reflective and controlled approach requires effort and time. Importantly, higher levels are grounded in the lower level control system and the overall system showed both kinds of behavior: On the one hand, the system showed again to be robust and adaptive as stable gait patterns emerged.

On the other hand, when the system was forced into an unstable situation, the cognitive extension intervened and found successful solutions through recruiting lower level behaviors out of context on a longer timescale.

Therefore, the system realizes a minimal cognitive system as it shows two modes of operation. In the introduction, we introduced a dual process view and the presented system fits to the two types of processes distinguished. This architecture is still minimal. In particular with respect to type 2 (higher level) processes, the realized function matches qualitatively posed requirements (controlled processing, effortful, reflective, decoupling and a form of mental simulation), but treatments of such type 2 processes are often more far reaching and include much more higher level knowledge (abstract and domain general knowledge). On the one hand, we purposefully focussed on such a very restricted set of behavior that is accessible on different levels of inquiry. This allows to come up with models that are broadly grounded in experimental findings and provide a realization on a computational and algorithmic level (Marr, 1982). Such minimal cognitive systems offer ideal tools for further investigation as their structure is interpretable and their function can be analyzed in detail (which is further helped by the modular and hierarchical structure). As one example, we analyzed in detail how and what kind of higher level mental states emerged in our architecture (Cruse & Schilling, 2015*a*,*b*). It is important to note that relating such an architecture towards attention, intention, or emotion was not a goal for producing such an architecture. But, interestingly, characteristics of emerging mental states match well functional descriptions of these higher level mental states. The organizing principles of our system appear sufficient to explain, on a functional level, the emergence of such mental states (For further details see the original publications (Cruse & Schilling, 2015*a*,*b*, 2013). Furthermore, we add a brief discussion article of higher mental states in the appendix (Schilling & Cruse, 2016, see E.1, page 310). This article was a reply towards a perspective that connected description of higher mental capabilities to insects. As a main point, we advised caution in interpretation of such findings. From our point of view, insects—and animals in general—provide ideal models for quantitative analysis of mechanisms on a functional and an algorithmic level. But such models should then only be used in this way. One should refrain from relating models to purely descriptive approaches, but should try to offer functional application that will be much more clear-cut and, as found in our case, can provide a clear decomposition in characteristics or key principles.).

On the other hand, the minimal cognitive system is minimalistic with respect to internal representation and how these support cognitive function. While planning ahead is realized as internal simulation, this only shows one specific example of recruitment of grounded internal models. A key advantage of the notion of recruitment is that it is, first, parsimonious to reuse models for different cognitive function, and, secondly, that concepts are actually constituted in their connections to different modalities. This idea found support in the findings of mirror neurons (Rizzolatti *et al.*, 1996) and the mirror neuron system (Cook *et al.*, 2014; Rizzolatti, 2005) which shows activation of motor control related areas in the brain during observation of actions. We are currently working towards extending our system in order to recruit the existing internal body model during observation. As knowledge on the body structure and the dynamics of possible movements is encoded inside the internal body model, this can be used as a prior during observation of movements and can guide—without the need for explicit supervision—finding correlations between visual inputs during observation and own observed movements (A simplified example was already used as a first proof of concept and is added in the appendix (Schilling, 2011*a*, E.2, page 312). It addresses findings in humans that show an advantage in an observational task for exploiting their own internal model (Loula *et al.*, 2005).).

Currently, our model does not address abstract or domain general knowledge. It is future work to extend the existing internal models to integrate, first, knowledge on distal information. Importantly, this should follow an embodied perspective. Further information should be incorporated in relation to the system itself, i.e., what something affords to the system (Gibson, 1977, 1979). One straight forward example is navigation as found in insects as well (Hoinville & Wehner, 2018). This might ground a spatial organization for conceptual spaces as found in animals (Moser *et al.*, 2008; Burgess, 2014). Spatial representations appear as one key conceptual space that is related to bodily representations (Romano *et al.*, 2017). It is furthermore involved in a form of mental simulation as replay activation of spatial representations corresponds to imagination of possible choices in navigational tasks (Ólafsdóttir *et al.*, 2015; Wu *et al.*, 2017).

Second, in our approach the cognitive expansion is only concerned with the selection of a single behavior out of context. As future work, we are interested in planning complex behaviors that consist of concurrent and sequential activation of behaviors. This requires a form of compositionality (Lake *et al.*, 2017; Cangelosi *et al.*, 2010). Following our bottom-up approach, we are interested in the structure of action representation for more complex behaviors. As one key characteristic, such action representations appear schematic (Binder & Desai, 2011). This has been well studied for a long time from a high level perspective in linguistics (Johnson, 1987; Fillmore, 1976) and has converged now with the idea of recruitment that language recruits such underlying schematic action representations (Pulvermüller, 2018; Gallese & Lakoff, 2005). There is now more and more evidence showing activation of neural areas associated with motor control that are differentially activated in language processing and imagination of semantics (Wang *et al.*, 2018; Desai *et al.*, 2013; Boulenger *et al.*, 2008) (Recently, we discussed the schematic organization and how this is recruited in language in an article which for further details is part of the appendix (Schilling *et al.*, 2020, E.3, page 322). Furthermore, we laid out a principal connection of our bottom-up approach to such a high level schematic view as a process model in (Schilling & Narayanan, 2013).).

One important capability to extend our system—in particular with respect to expanding information that can be used by the system—is learning. Current work—presented in the third chapter—addresses already Deep Reinforcement Learning on the lower levels. The goal is to extend this towards more complex actions and towards a broader range of behaviors, for example, using the two front legs for grasping objects and moving them around in an environment. This will require to introduce more hierarchical structures and different temporal scales into the learning framework. For the higher levels, Deep Reinforcement Learning (and in particular hierarchical DRL) appears as well as a natural extension to our system, as we are already realizing a simple form of trial-and-error learning on the action selection level during cognitive processing. While this would change (or even replace) the structure of the out-of-context action selection neural network, such an approach would benefit from the particular characteristics that would be maintained:

- a modular local structure on the lower level that allows for fast adaptations,
- decentralized action selection that is coordinated through local rules,
- and exploiting predictive capabilities of an internal body model for planning actions to deal with novel situations (we have proposed such a formulation of a decentralized learning system in Schilling & Melnik (2018)).

Such a system could provide a basis for analysis of interactions between different levels of decentralized concurrent control that integrates internal models and could be used for learning on longer timescales.

6.3 List of Publications

As this chapter gave an outlook on further extensions and application of the proposed model, there are three articles that provide further discussion on the process of mental simulation and the structure of underlying representation.

6.3.1 Contributions to the Thesis

- Schilling, M., and Cruse, H. (2016), "Avoid the hard problem: Employment of mental simulation for prediction is already a crucial step".

  Appendix E.1, page 310: Discussion on importance of mental simulation as a central mechanism, and relating this to findings in insects. This was published in Proceedings of the National Academy of Sciences (PNAS, Impact Factor 9.58).
  Author Contributions: M.S. and H.C. devised the main conceptual ideas, laid out the concept and wrote the paper.

- Schilling, M. (2011), "Learning by seeing—associative learning of visual features through mental simulation of observed action". Proceedings of the European Conference on Artificial Life 2011, Paris.

  Appendix E.2, page 312: Provides a proof of concept for recruitment of the internal body model in an observational task.

- Schilling, M., Chang, N., Rohlfing, K.J., Spranger, M. (in press), "Simulation across Representation: The Interplay of Schemas and Simulation-Based Inference on Different Levels of Abstraction".

  Appendix E.3, page 322: Discussion on the schematic structure of action representation from a neuroscientific and linguistic perspective. This was recently accepted for publication in the journal Behavioral and Brain Sciences (Impact Factor 15.07).
  Author Contributions: M.S., N.C., K.J.R., and M.Sp. devised the project, the main conceptual ideas, and wrote the paper. M.S. prepared the original draft of the paper.

6.3.2 Further Related Publications

There are further related publications that relate to the structure of representation, higher level function in cognitive systems, and recently turning towards interaction as well as interactive systems.

Peer-reviewed Journal Papers:

- Schilling, M., Burgard. W., Muelling, K., Wrede, B. and Ritter, H. (2019), "Shared Autonomy— Learning of Joint Action and Human-Robot Collaboration". Frontiers in Neurorobotics 13:16. doi:10.3389/fnbot.2019.00016

- Nomikou, I., Schilling, M., Heller, V. and Rohlfing, K. J. (2016), "Language at all times. Action and interaction as contexts for enriching representations". Interaction Studies, 17(1), pp. 128–153.

- Cruse, H. and Schilling, M. (2013), "How and to what end may consciousness contribute to action? Attributing properties of consciousness to an embodied, minimally cognitive artificial neural network". Frontiers in Psychology, 4(324). doi: 10.3389/fpsyg.2013.00324

Book Chapters:

- Cruse, H. and Schilling, M. (2016), "Mental states as emergent properties. From walking to consciousness". In: Open Mind, Philosophy and the Mind Sciences in the 21st Century. Vol 1. Metzinger T, Windt JM (Eds); Cambridge, Mass.: The MIT Press: 349-386.

- Schilling, M. (2012), "Grounded internal body models for communication: Integration of sensory, motor and visual spaces for mediating conceptualization". In L. Steels and M. Hild (Eds.), Language Grounding in Robots. Berlin: Springer (pages 131–150).

Reviewed Conference Proceedings:

- Schilling, M., Kopp, S., Wachsmuth, S., Wrede, B., Ritter, H., Brox, T., Nebel, B., Burgard, W. (2016). “Towards A Multidimensional Perspective on Shared Autonomy”. Proceedings of the AAAI Fall Symposium Series 2016, Stanford (USA).

- Schilling, M. and Narayanan, S. (2013), “Communicating with Executable Action Representations”. In Proceedings of AAAI Spring Symposium Series 2013, Stanford.

- Cruse, H., and Schilling, M. (2011), “From egocentric systems to systems allowing for theory of mind and mutualism”. In R. Doursat (Ed.), Proceedings of the ECAL 2011, Paris: MIT Press, pp. 184–191.

- Schilling, M. (2011), “Integrating multi-sensory input in the body model – a RNN approach to connect proprioception, visual features and motor control”. Proc. of the International Joint Conference on Neural Networks 2011, San Jose (CA).

Appendices

Overview Articles

This habilitation thesis presents the author's selected findings on the organization, realization, and learning of motor control structures and cognitive function applied on a six-legged walking robot. It is written as a cumulative thesis. This appendix contains the main scientific contributions that have been published in 15 papers. While two articles are submitted, all other articles have been peer-reviewed. Seven articles have been published as a journal publication (mean impact factor of 5.65) and the other six have been published at international conferences (either high ranked CORE A conferences or specialized conferences on specific topics). These articles constitute the second and major part of the thesis—further related publications are pointed out in the chapters as well.

Appendix A – Biological-inspired Locomotion Control

- Schilling, M., Hoinville, T., Schmitz, J. and Cruse, H. (2013), "Walknet, a bio-inspired controller for hexapod walking". Biological Cybernetics, 107(4), pages 397–419.

  Appendix A.1, page 94: Provides a review on behavioral findings in insects and details on the Walknet control approach (Schilling *et al.*, 2013*a*, A.1, page 94). This was published in Biological Cybernetics (Impact Factor 1.76) and is now established as a reference for decentralized organization of motor control in insects (cited 129 times).
  Author Contributions: MS and HC laid out the concept, designed the model and analyzed the data. MS carried out the implementation. MS and HC wrote the manuscript. Writing review and editing all authors.

- Schilling, M. and Cruse, H. (2020), "Decentralized control of insect walking - a simple neural network explains a wide range of behavioral and neurophysiological results". PLOS Computational Biology 16(4): e1007804. https://doi.org/10.1371/journal.pcbi.1007804

  Appendix A.2, page 118: Extension towards a detailed decentralized architecture that acts on the joint level. This was recently accepted for publication in PLOS Computational Biology (Impact Factor 4.43).
  Author Contributions: Conceptualization, methodology, investigation, and writing – MS and HC. Software, simulation, and data curation – MS. Formal analysis – HC.

- Schilling, M., Paskarbeit, J., Hüffmeier, A., Schneider, A., Schmitz, J., and Cruse, H. (2013), "A hexapod walker using a heterarchical architecture for action selection". Frontiers in Computational Neuroscience 7:126. doi: 10.3389/fncom.2013.00126.

  Appendix A.3, page 168: Introduces the hierarchical Motivation Unit architecture and shows results for application on a hexapod robot. The article was published in Frontiers in Computational Neuroscience (Impact Factor 3.57). Author Contributions: Conceptualization and writing – MS and HC. Methodology and investigation – MS, HC, JS, and AS. Software and simulation – JP, AH, and MS.

Appendix B – Decentralized Learning

- Schilling, M., Ritter, H., and Ohl, F.W. (2019), "From Crystallized Adaptivity to Fluid Adaptivity in Deep Reinforcement Learning — Insights from Biological Systems on Adaptive Flexibility". In 2019 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Bari (I).

  Appendix B.1, page 188: Discusses in detail learning on different timescales as a form of fluid adaptivity that allows to continuously adapt to changing environments and contrasts this with approaches that learn how to switch between different types of behaviors. Introduces deceptive problems in more detail. This was published at IEEE SMC (ranking: CORE B).
  Author Contributions: Conceptualization and writing original draft – MS. Investigation, writing review and editing – MS, FWO, HR.

- Schilling, M., Konen, K., Ohl, F.W., and Korthals, T. (submitted to IROS conference), "Decentralized Deep Reinforcement Learning for a Distributed and Adaptive Locomotion Controller of a Hexapod Robot".

  Appendix B.2, page 196: Application of decentralized controllers in a DRL setting. Showing detailed results that such a decentralized architecture learns robust controllers that even perform better compared to a baseline approach.
  Author Contributions: Conceptualization, writing original draft and data curation – MS. Methodology, formal analysis – KK, TK, MS. Investigation and software – KK. Writing review and editing – MS, KK, FWO, TK.

- Schilling, M. and Melnik, A. (2018), "An Approach to Hierarchical Deep Reinforcement Learning for a Decentralized Walking Control Architecture". In: Samsonovich A. (eds.) Biologically Inspired Cognitive Architectures 2018. Advances in Intelligent Systems and Computing, vol 848. Springer, Cham.

  Appendix B.3, page 206: Introduces how to extend the decentralized architecture towards a hierarchical approach in a DRL setting (implementation is future work).
  Author Contributions: Conceptualization and writing original draft – MS. Methodology, investigation, writing review and editing – MS, AM.

Appendix C – Hierarchical Internal Model

- Schilling, M., Paskarbeit, J., Schmitz, J., Schneider, A., and Cruse, H. (2012), "Grounding an Internal Body Model of a Hexapod Walker—Control of Curve Walking in a Biological Inspired Robot". In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, pages 2762-2768.

  Appendix C.1, page 220: Details on the setup and processing of the hierarchical internal body model for the case of a six-legged robot. Shown are results for the control of the stance movement in forward and curve walking in simulation. This was published in the main track of the IROS conference, one of the three leading conferences in the field of robotics (CORE A ranking).
  Author Contributions: Conceptualization, methodology, investigation, analysis – MS. Software, simulation – MS, JP. Writing – MS, JP, JS, AS, HC.

- Schilling, M., and Cruse, H. (2012), "What's next: Recruitment of a grounded predictive body model for planning a robot's actions". Frontiers in Cognition, 3(383). doi:10.3389/fpsyg.2012.00383

  Appendix C.2, page 228: Provides an introduction into the MMC principle and shows predictive capabilities of this model and coordination in a targeted reaching task for an insect-like body structure. This was published in Frontiers in Cognition (Impact Factor 2.13).
  Author Contributions: Conceptualization, methodology, software, simulation, investigation, and writing – MS. Writing, review and editing as well as funding acquisition – HC.

- Schilling, M. (2019), "Hierarchical Dual Quaternion-Based Recurrent Neural Network as a Flexible Internal Body Model". Proc. of the International Joint Conference on Neural Networks 2019, Budapest (Hungary), pp. 1–8.

  Appendix C.3, page 248: A hierarchical body model for more complex manipulators. Introduces the dual quaternion representation for joints with multiple degrees of freedom and the extension towards a hierarchical model. Shows results for a series of simulations for a simple bimanual task that required computation of forward and inverse kinematics. This was published at IJCNN (ranking: CORE A).

Appendix D – Planning Ahead in a Cognitive Architecture

- Schilling, M., Paskarbeit, J., Ritter, H., Schneider, A., and Cruse, H. (submitted), "From Adaptive Locomotion to Predictive Action Selection—Cognitive Control for a Six-Legged Walker".

  Appendix D.1, page 260: Summary of the structure of the cognitive expansion and application on the real robot Hector in a climbing task (Schilling *et al.*, submitted, 2020*b*, D.1, page 260). In addition, provides a systematic analysis

on postural variations and shows how adaptive system and cognitive system complement each other.
Author Contributions: M.S. and H.C. designed and performed research as well as analyzed the data. M.S. wrote simulation software and developed concept of the model. Simulation experiments were performed by M.S. and experiments on robot were performed by J.P and M.S. M.S. wrote the paper and H.C., H.R., and A.S. contributed to the writing.

- Schilling, M. and Cruse, H. (2017), "ReaCog, a Minimal Cognitive Controller Based on Recruitment of Reactive Systems". Front. Neurorobot. 11(3). doi: 10.3389/fnbot.2017.00003

  Appendix D.2, page 276: Introduces the cognitive expansion and demonstrates the application for the case of an awkward posture on a simulated robot (Schilling *et al.*, submitted, 2020*b*, D.2, page 276). This was published in Frontiers in Neurorobotics (Impact Factor 3.00).
  Author Contributions: MS and HC laid out the concept, designed the model and analyzed the data. MS carried out the implementation. MS and HC wrote the manuscript. Writing review and editing all authors.

- Schilling, M. (2017), "Old Actions in Novel Contexts — a Cognitive Architecture for Safe Explorative Action Selection". Proceedings of the Artificial Intelligence and Simulation of Behaviour Conference (AISB 2017), Bath (UK).

  Appendix D.3, page 300: Analyzes in detail the stages coordinating cognitive processing and shows their convergence (Schilling, 2017, see article D.3, page 300).

Appendix E – Further Perspective Articles

- Schilling, M., and Cruse, H. (2016), "Avoid the hard problem: Employment of mental simulation for prediction is already a crucial step".

  Appendix E.1, page 310: Discussion on importance of mental simulation as a central mechanism, and relating this to findings in insects. This was published in Proceedings of the National Academy of Sciences (PNAS, Impact Factor 9.58).
  Author Contributions: M.S. and H.C. devised the main conceptual ideas, laid out the concept and wrote the paper.

- Schilling, M. (2011), "Learning by seeing—associative learning of visual features through mental simulation of observed action". Proceedings of the European Conference on Artificial Life 2011, Paris.

  Appendix E.2, page 312: Provides a proof of concept for recruitment of the internal body model in an observational task.

- Schilling, M., Chang, N., Rohlfing, K.J., Spranger, M. (in press), "Simulation across Representation: The Interplay of Schemas and Simulation-Based Inference on Different Levels of Abstraction".

  Appendix E.3, page 322: Discussion on the schematic structure of action representation from a neuroscientific and linguistic perspective. This was recently accepted for publication in the journal Behavioral and Brain Sciences (Impact Factor 15.07).
  Author Contributions: M.S., N.C., K.J.R., and M.Sp. devised the project, the main conceptual ideas, and wrote the paper. M.S. prepared the original draft of the paper.

Bibliography

Ache, J. M. & Matheson, T. (2013), Passive Joint Forces Are Tuned to Limb Use in Insects and Drive Movements without Motor Activity. *Current Biology* **23** (15), 1418–1426.

Acosta-Calderon, C. & Hu, H. (2005), Robot imitation: body schema and body percept. *Journal of Applied Bionics and Biomechanics* **2** (3-4), 131–148.

Aljalbout, E., Walter, F., Röhrbein, F. & Knoll, A. (2020), Task-Independent Spiking Central Pattern Generator: A Learning-Based Approach. *Neural Processing Letters* .

Alon, U. (2006), *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Taylor & Francis.

Andersen, R. & Mountcastle, V. (1983), The influence of the angle of gaze upon the excitability of the light- sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience* **3** (3), 532–548.

Anderson, M. L. (2010), Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* **33**, 254–313.

Arber, S. & Costa, R. M. (2018), Connecting neuronal circuits for movement. *Science* **360** (6396), 1403–1404.

Arcaro, M. J., Schade, P. F. & Livingstone, M. S. (2019), Body map proto-organization in newborn macaques. *Proceedings of the National Academy of Sciences* **116** (49), 24861–24871.

Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. (2017), A Brief Survey of Deep Reinforcement Learning. *IEEE Signal Processing Magazine* **34** (6), 26–38.

Barsalou, L. W. (2008), Grounded Cognition. *Annual Review of Psychology* **59**, 617–645.

Baum, M., Meier, M. & Schilling, M. (2015), Population based mean of multiple computations networks: A building block for kinematic models. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE.

Beer, R. D. (1990), *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology*. Academic Press Professional, Inc., USA.

Beer, R. D., Chiel, H. J. & Sterling, L. S. (1990), A biological perspective on autonomous agent design. *Robotics and Autonomous Systems* **6** (1), 169 – 186. Designing Autonomous Agents.

Bellicoso, C. D., Jenelten, F., Gehring, C. & Hutter, M. (2018), Dynamic Locomotion Through Online Nonlinear Motion Optimization for Quadrupedal Robots. *IEEE Robotics and Automation Letters* **3** (3), 2261–2268.

Bernstein, N. A. (1967), *The Co-ordination and regulation of movements*. Pergamon Press Ltd., Oxford.

Bernstein, N. A. (1996), On dexterity and its development. In *Dexterity and its Development* (ed. M. Latash & M. Turvey), pp. 1–244. L. Erlbaum Associates.

Bidaye, S. S., Bockemühl, T. & Büschges, A. (2018), Six-legged walking in insects: how CPGs, peripheral feedback, and descending signals generate coordinated and adaptive motor rhythms. *Journal of Neurophysiology* **119** (2), 459–475.

Billard, A. & Kragic, D. (2019), Trends and challenges in robot manipulation. *Science* **364** (6446), eaat8414.

Binder, J. R. & Desai, R. H. (2011), The neurobiology of semantic memory. *Trends in Cognitive Sciences* **15** (11), 527–536.

Binder, M. D., Hirokawa, N. & Windhorst, U. (2009), Motor Control Hierarchy. In *Encyclopedia of Neuroscience*, pp. 2428–2428. Springer, Berlin, Heidelberg.

Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. (1991), Computations underlying the execution of movement: a biological perspective. *Science* **253** (5017), 287–291.

Bongard, J. & Lipson, H. (2014), Evolved Machines Shed Light on Robustness and Resilience. *Proceedings of the IEEE* **102** (5), 899–914.

Bongard, J., Zykov, V. & Lipson, H. (2006), Resilient Machines Through Continuous Self-Modeling. *Science* **314** (5802), 1118–1121.

Botvinick, M., Barrett, D. G., Battaglia, P., de Freitas, N., Kumaran, D., Leibo, J. Z., Lillicrap, T., Modayil, J., Mohamed, S., Rabinowitz, N. C. *et al.* (2017), Building machines that learn and think for themselves. *Behavioral and Brain Sciences* **40**.

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C. & Hassabis, D. (2019), Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences* **23** (5), 408–422.

Botvinick, M. M. (2007), Multilevel structure in behaviour and in the brain: a model of Fuster's hierarchy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **362** (1485), 1615–26.

Botvinick, M. M. (2008), Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences* **12** (5), 201–208.

Boulenger, V., Hauk, O. & Pulvermüller, F. (2008), Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. *Cerebral Cortex* **19** (8), 1905–1914.

Brette, R. (2019), Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences* **42**.

Broekens, J. (2005), Internal Simulation of Behavior has an Adaptive Advantage. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, pp. 342–347. Lawrence Erlbaum Associates.

Brooks, R. (1986), A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation* **2** (1), 14–23.

Brooks, R. A. (1989), A robot that walks; emergent behaviors from a carefully evolved network. *Neural Computation* **1** (2), 253–262.

Brooks, R. A. (1991), Intelligence Without Reason. In *Proceedings of the 12th international joint conference on Artificial intelligence* (ed. J. Myopoulos & R. Reiter), pp. 569–595. Morgan Kaufmann Publishers Inc.: San Mateo, CA, USA.

Brown, I. E. & Loeb, G. E. (2000), A reductionist approach to creating and using neuromusculoskeletal models. In *Biomechanics and neural control of posture and movement*, pp. 148–163. Springer.

Burgess, N. (2014), The 2014 Nobel Prize in Physiology or Medicine: a spatial model for cognitive neuroscience. *Neuron* **84** (6), 1120–5.

Büschges, A., Schmitz, J. & Bässler, U. (1995), Rhythmic patterns in the thoracic nerve cord of the stick insect induced by pilocarpine. *Journal of Experimental Biology* **198** (2), 435–456.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., Tani, J., Belpaeme, T., Sandini, G. & Nori, F. (2010), Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development* **2** (3), 167–195.

Card, G. & Dickinson, M. H. (2008), Visually Mediated Motor Planning in the Escape Response of Drosophila. *Current Biology* **18**, 1300–1307.

Carlo, J. D., Wensing, P. M., Katz, B., Bledt, G. & Kim, S. (2018), Dynamic Locomotion in the MIT Cheetah 3 Through Convex Model-Predictive Control. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9.

Chersi, F., Donnarumma, F. & Pezzulo, G. (2013), Mental imagery in the navigation domain: a computational model of sensory-motor simulation mechanisms. *Adaptive Behavior* **21** (4), 251–262.

Chiel, H. J. & Beer, R. D. (1997), The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences* **20** (12), 553–557.

Chiel, H. J., Ting, L. H., Ekeberg, O. & Hartmann, M. J. Z. (2009), The Brain in Its Body: Motor Control and Sensing in a Biomechanical Context. *Journal of Neuroscience* **29** (41), 12807–12814.

Clancy, K. B., Orsolic, I. & Mrsic-Flogel, T. D. (2019), Locomotion-dependent remapping of distributed cortical networks. *Nature Neuroscience* **22** (5), 778.

Clune, J., Mouret, J.-B. & Lipson, H. (2013), The evolutionary origins of modularity. *Proceedings of the Royal Society B: Biological Sciences* **280** (1755), 20122863.

Colas, C., Sigaud, O. & Oudeyer, P. (2019), A hitchhiker's guide to statistical comparisons of reinforcement learning algorithms. In *Reproducibility in Machine Learning, ICLR 2019 Workshop, New Orleans (US).*

Cook, R., Bird, G., Catmur, C., Press, C. & Heyes, C. (2014), Mirror neurons: from origin to function. *Behavioral and Brain Sciences* **37** (2), 177–92.

Cruse, H. (1979), The control of the anterior extreme position of the hindleg of a walking insect. *Physiological Entomology* **4**, 121–124.

Cruse, H. (1999), Feeling our body – the basis of cognition? *Evolution and Cognition* **5** (2), 162–173.

Cruse, H. & Schilling, M. (2013), How and to what end may consciousness contribute to action? Attributing properties of consciousness to an embodied, minimally cognitive artificial neural network. *Frontiers in Psychology* **4**, 324.

Cruse, H. & Schilling, M. (2015*a*), The Bottom-up Approach: Benefits and Limits. In *Open Mind* (ed. T. K. Metzinger & J. M. Windt). MIND Group.

Cruse, H. & Schilling, M. (2015*b*), Mental states as emergent properties. From walking to consciousness. In *Open Mind* (ed. T. Metzinger & J. Windt). Frankfurt am Main: MIND Group.

Cruse, H., Steinkühler, U. & Burkamp, C. (1998), MMC - a recurrent neural network which can be used as manipulable body model. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior* (ed. R. Pfeifer, B. Blumberg, J.-A. Meyer & S. W. Wilson), pp. 381–389.

Cully, A., Clune, J., Tarapore, D. & Mouret, J.-B. (2015), Robots that can adapt like animals. *Nature* **521** (7553), 503–507.

d'Avella, A., Giese, M., Ivanenko, Y. P., Schack, T. & Flash, T. (2015), Editorial: Modularity in motor control: from muscle synergies to cognitive action representation. *Frontiers in Computational Neuroscience* **9**, 126.

Daw, N. D., Niv, Y. & Dayan, P. (2005), Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* **8** (12), 1704–1711.

DeAngelis, B. D., Zavatone-Veth, J. A. & Clark, D. A. (2019), The manifold structure of limb coordination in walking Drosophila. *eLife* **8**, e46409.

Desai, R. H., Conant, L. L., Binder, J. R., Park, H. & Seidenberg, M. S. (2013), A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage* **83**, 862–869.

Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S., Wu, Y. & Zhokhov, P. (2017), OpenAI Baselines. `https://github.com/openai/baselines`.

Dickinson, A. (1985), Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **308** (1135), 67–78.

Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M. a. R., Kram, R. & Lehman, S. (2000), How Animals Move: An Integrative View. *Science* **288** (5463), 100–106.

Dudek, D. M. & Full, R. J. (2006), Passive mechanical properties of legs from running insects. *The Journal of Experimental Biology* **209** (Pt 8), 1502–1515.

Dürr, V., Arena, P., Cruse, H., Dallmann, C. J., Drimus, A., Hoinville, T., Krause, T., Mátéfi-Tempfli, S., Paskarbeit, J., Patanè, L., Schäffersmann, M., Schilling, M., Schmitz, J., Strauss, R., Theunissen, L., Vitanza, A. & Schneider, A. (2019), Integrative Biomimetics of Autonomous Hexapedal Locomotion. *Frontiers in Neurorobotics* **13**.

Dürr, V. & Ebeling, W. (2005), The behavioural transition from straight to curve walking: kinetics of leg movement parameters and the initiation of turning. *Journal of Experimental Biology* **208** (12), 2237–2252.

Dürr, V. & Schilling, M. (2018), Transfer of Spatial Contact Information Among Limbs and the Notion of Peripersonal Space in Insects. *Frontiers in Computational Neuroscience* **12**.

Dürr, V., Schmitz, J. & Cruse, H. (2004), Behaviour-based modelling of hexapod locomotion: Linking biology and technical application. *Arthropod Structure and Development* **33** (3), 237–250.

Ellefsen, K. O., Huizinga, J. & Torresen, J. (2020), Guiding Neuroevolution with Structural Objectives. *Evolutionary Computation* **28** (1), 115–140.

Evans, J. S. B. T. (2008), Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology* **59** (1), 255–278.

Evans, J. S. B. T. & Stanovich, K. E. (2013), Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science* **8** (3), 223–241. PMID: 26172965.

Fazeli, N., Oller, M., Wu, J., Wu, Z., Tenenbaum, J. B. & Rodriguez, A. (2019), See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion. *Science Robotics* **4** (26), eaav3123.

Fillmore, C. J. (1976), Frame semantics and the nature of language. *Annals of the New York Academy of Sciences* **280** (1), 20–32.

Finn, C., Abbeel, P. & Levine, S. (2017), Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *International Conference on Machine Learning*, pp. 1126–1135.

Flash, T. & Hochner, B. (2005), Motor primitives in vertebrates and invertebrates. *Current Opinion in Neurobiology* **15** (6), 660–6.

Foerster, J. N., Farquhar, G., Afouras, T., Nardelli, N. & Whiteson, S. (2018), Counterfactual multi-agent policy gradients. In *Thirty-second AAAI conference on artificial intelligence.*

Frans, K., Ho, J., Chen, X., Abbeel, P. & Schulman, J. (2018), Meta Learning Shared Hierarchies. In *International Conference on Learning Representations.*

Fujimoto, S., van Hoof, H. & Meger, D. (2018), Addressing function approximation error in actor-critic methods. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden* (ed. J. G. Dy & A. Krause), *Proceedings of Machine Learning Research*, vol. 80, pp. 1582–1591. PMLR.

Full, R. J. & Tu, M. S. (1991), Mechanics of a rapid running insect: two-, four- and six-legged locomotion. *Journal of Experimental Biology* **156** (1), 215–231.

Gallese, V. & Lakoff, G. (2005), The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* **22** (3-4), 455–479.

Georgopoulos, A., Kettner, R. & Schwartz, A. (1988), Primate motor cortex and free arm movements to visual targets in three- dimensional space. II. Coding of the direction of movement by a neuronal population. *Journal of Neuroscience* **8** (8), 2928–2937.

Gibson, J. J. (1977), The Theory of Affordances. In *Perceiving, Acting, and Knowing* (ed. J. B. Robert Shaw), pp. 67–80. Lawrence Erlbaum Associates.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception.* New Jersey: Lawrence Erlbaum Associates.

Giszter, S. F., Mussa-Ivaldi, F. A. & Bizzi, E. (1993), Convergent force fields organized in the frog's spinal cord. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **13** (2), 467–491.

Giurfa, M. & Menzel, R. (2013), Cognitive Components of Insect Behavior. In *Handbook of Behavioral Neuroscience 22*, pp. 14–25. Elsevier.

Glenberg, A. M. (1997), What memory is for. *Behavioral and Brain Sciences* **20** (1), 1–55.

Glenberg, A. M. & Gallese, V. (2012), Action-based language: A theory of language acquisition, comprehension, and production. *Cortex* **48** (7), 905–922.

Graham, D. (1972), A behavioural analysis of the temporal organisation of walking movements in the 1st instar and adult stick insect (Carausius morosus). *Journal of Comparative Physiology* **81** (1), 23–52.

Graziano, M. (2006), The Organization of Behavioral Repertoire in Motor Cortex. *Annual Review of Neuroscience* **29**, 105–134.

Grillner, S. (2003), The motor infrastructure: from ion channels to neuronal networks. *Nature Reviews Neuroscience* **4** (7), 573–86.

Grillner, S. & Zangger, P. (1979), On the central generation of locomotion in the low spinal cat. *Experimental Brain Research* **34** (2), 241–261.

Han, D., Doya, K. & Tani, J. (2020), Self-organization of action hierarchy and compositionality by reinforcement learning with recurrent neural networks. *Neural Networks* (in press).

Hart, C. B. & Giszter, S. F. (2010), A Neural Basis for Motor Primitives in the Spinal Cord. *Journal of Neuroscience* **30** (4), 1322–1336.

Haruno, M., Wolpert, D. M. & Kawato, M. (2003), Hierarchical MOSAIC for movement generation. *International Congress Series* **1250**, 575–590.

Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. (2017), Neuroscience-Inspired Artificial Intelligence. *Neuron* **95** (2), 245–258.

Hauser, H. (2019), Resilient machines through adaptive morphology. *Nature Machine Intelligence* pp. 1–2.

Hay, N., Stark, M., Schlegel, A., Wendelken, C., Park, D., Purdy, E., Silver, T., Phoenix, D. S. & George, D. (2018), Behavior is everything: Towards representing concepts with sensorimotor contingencies. In *AAAI Conference on Artificial Intelligence*.

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. *et al.* (2017), Emergence of locomotion behaviours in rich environments. *arXiv preprint arXiv:1707.02286* .

Heess, N., Wayne, G., Tassa, Y., Lillicrap, T., Riedmiller, M. & Silver, D. (2016), Learning and transfer of modulated locomotor controllers. *arXiv preprint arXiv:1610.05182* .

Hesslow, G. (2002), Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences* **6** (6), 242–247.

Hoffmann, M., Marques, H., Arieta, A. H., Sumioka, H., Lungarella, M. & Pfeifer, R. (2010), Body schema in robotics: a review. *IEEE Trans. Auton. Mental Develop.* **2** (4), 304–324.

Hoinville, T. & Wehner, R. (2018), Optimal multiguidance integration in insect navigation. *Proceedings of the National Academy of Sciences* **115** (11), 2824–2829.

Honegger, H.-W. (1981), A preliminary note on a new optomotor response in crickets: Antennal tracking of moving targets. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* **142** (3), 419–421.

Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V. & Hutter, M. (2019), Learning agile and dynamic motor skills for legged robots. *Science Robotics* **4** (26), eaau5872.

Ijspeert, A. J. (2008), Central pattern generators for locomotion control in animals and robots: a review. *Neural Networks* **21** (4), 642–53.

Ijspeert, A. J. (2014), Biorobotics: using robots to emulate and investigate agile locomotion. *Science* **346** (6206), 196–203.

Ijspeert, A. J. (2018), Decoding the neural mechanisms underlying locomotion using mathematical models and bio-inspired robots: From lamprey to human locomotion. In *Robotics Research: Volume 1* (ed. A. Bicchi & W. Burgard), pp. 177–186. Springer International Publishing, Cham.

Ikeda, K. & Wiersma, C. (1964), Autogenic rhythmicity in the abdominal ganglia of the crayfish: the control of swimmeret movements. *Comparative biochemistry and physiology* **12** (1), 107–115.

Jindrich, D. L. & Full, R. J. (2002), Dynamic stabilization of rapid hexapedal locomotion. *The Journal of Experimental Biology* **205** (Pt 18), 2803–2823.

Johnson, M. (1987), *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Kahneman, D. (2011), *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.

Kahneman, D. & Frederick, S. (2002), Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In *Heuristics and Biases*, 1st edn. (ed. T. Gilovich, D. Griffin & D. Kahneman), pp. 49–81. Cambridge University Press.

Kalakrishnan, M., Buchli, J., Pastor, P., Mistry, M. & Schaal, S. (2010), Fast, robust quadruped locomotion over challenging terrain. In *2010 IEEE International Conference on Robotics and Automation*, pp. 2665–2670.

Kashtan, N. & Alon, U. (2005), Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences* **102** (39), 13773–13778.

Kim, S. & Wensing, P. M. (2017), Design of Dynamic Legged Robots. *Found. Trends Robot* **5** (2), 117–190.

Knebel, D., Ayali, A., Pflüger, H.-J. & Rillich, J. (2017), Rigidity and flexibility: The central basis of inter-leg coordination in the locust. *Frontiers in Neural Circuits* **10**, 112.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. (2017), Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93** (3), 480–490.

Kulkarni, T., Narasimhan, K., Saeedi, A. & Tenenbaum, J. B. (2016), Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In *Advances in Neural Information Processing Systems*, pp. 3675–3683.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. (2017), Building machines that learn and think like people. *Behavioral and brain sciences* **40**.

Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Pérolat, J., Silver, D. & Graepel, T. (2017), A unified game-theoretic approach to multiagent reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4193–4206. Curran Associates Inc., Red Hook, NY, USA.

Latash, M. L., Scholz, J. P. & Schöner, G. (2007), Toward a New Theory of Motor Synergies. *Motor Control* **11**, 276–308.

LeCun, Y., Bengio, Y. & Hinton, G. (2015), Deep learning. *Nature* **521** (7553), 436–444.

Lee, J. H., Seymour, B., Leibo, J. Z., An, S. J. & Lee, S. W. (2019), Toward high-performance, memory-efficient, and fast reinforcement learning—Lessons from decision neuroscience. *Science Robotics* **4** (26), eaav2975.

Lehman, J. & Stanley, K. O. (2011), Abandoning objectives: evolution through the search for novelty alone. *Evolutionary Computation* **19** (2), 189–223.

Limongelli, L., Boysen, S. T. & Visalberghi, E. (1995), Comprehension of cause-effect relations in a tool-using task by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology* **109** (1), 18.

Lipson, H. (2019), Robots on the run. *Nature* **568** (7751), 174.

Lipson, H., Pollack, J. B. & Suh, N. P. (2002), On the origin of modular variation. *Evolution* **56** (8), 1549–1556.

Loukola, O. J., Perry, C. J., Coscos, L. & Chittka, L. (2017), Bumblebees show cognitive flexibility by improving on an observed complex behavior. *Science* **355** (6327), 833–836.

Loula, F., Prasad, S., Harber, K. & Shiffrar, M. (2005), Recognizing People From Their Movement. *Journal of Experimental Psychology: Human Perception and Performance* **31** (1), 210–220.

Lowe, R., WU, Y., Tamar, A., Harb, J., Pieter Abbeel, O. & Mordatch, I. (2017), Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems 30* (ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett), pp. 6379–6390. Curran Associates, Inc.

Magill, R. & Anderson, D. I. (2017), *Motor learning and control: concepts and applications*. McGraw-Hill Education.

Marr, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., USA.

McFarland, D. & Bösser, T. (1993), *Intelligent behavior in animals and robots*. MIT Press, Cambridge, MA.

McGeer, T. (1993), Dynamics and Control of Bipedal Locomotion. *Journal of theoretical Biology* **163**, 277–314.

McNamee, D. & Wolpert, D. M. (2019), Internal Models in Biological Control. *Annual Review of Control, Robotics, and Autonomous Systems* **2** (1), 339–364.

Mengistu, H., Huizinga, J., Mouret, J.-B. & Clune, J. (2016), The evolutionary origins of hierarchy. *PLOS Comput Biol* **12** (6), e1004829.

Menzel, R., Brembs, B. & Giurfa, M. (2007), Cognition in invertebrates. In *Evolution of Nervous Systems in Invertebrates – 2* (ed. J. Kaas), pp. 403–442. Academic Press, Oxford.

Merel, J., Botvinick, M. & Wayne, G. (2019*a*), Hierarchical motor control in mammals and machines. *Nature Communications* **10** (1), 1–12. 00000.

Merel, J., Tunyasuvunakool, S., Ahuja, A., Tassa, Y., Hasenclever, L., Pham, V., Erez, T., Wayne, G. & Heess, N. (2019*b*), Reusable neural skill embeddings for vision-guided whole body movement and object manipulation. *arXiv preprint arXiv:1911.06636* .

Miller, G. S. P. (1986), The definition and rendering of terrain maps. *SIGGRAPH Comput. Graph.* **20** (4), 39–48.

Mischiati, M., Lin, H.-T., Herold, P., Imler, E., Olberg, R. & Leonardo, A. (2015), Internal models direct dragonfly interception steering. *Nature* **517** (7534), 333–8.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. (2015), Human-level control through deep reinforcement learning. *Nature* **518** (7540), 529–533.

Montesano, L., Lopes, M., Bernardino, A. & Santos-Victor, J. (2008), Learning object affordances: from sensory–motor coordination to imitation. *IEEE Transactions on Robotics* **24** (1), 15–26.

Morasso, P., Casadio, M., Mohan, V., Rea, F. & Zenzeri, J. (2015), Revisiting the body-schema concept in the context of whole-body postural-focal dynamics. *Frontiers in Human Neuroscience* **9**, 83.

More, H. L. & Donelan, J. M. (2018), Scaling of sensorimotor delays in terrestrial mammals. *Proceedings of the Royal Society B: Biological Sciences* **285** (1885).

Moser, E. I., Kropff, E. & Moser, M.-B. (2008), Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience* **31**, 69–89.

Mountcastle, V. B. (1997), The columnar organization of the neocortex. *Brain* **120** (4), 701–722.

Mouret, J.-B. & Clune, J. (2015), Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909* .

Nachum, O., Gu, S. S., Lee, H. & Levine, S. (2018), Data-efficient hierarchical reinforcement learning. In *Advances in Neural Information Processing Systems 31* (ed. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett), pp. 3303–3313. Curran Associates, Inc.

Namikawa, J., Nishimoto, R. & Tani, J. (2011), A Neurodynamic Account of Spontaneous Behaviour. *PLoS Computational Biology* **7** (10), e1002221.

Neftci, E. O. & Averbeck, B. B. (2019), Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence* **1**, 133–143.

Nishikawa, K., Biewener, A. A., Aerts, P., Ahn, A. N., Chiel, H. J., Daley, M. A., Daniel, T. L., Full, R. J., Hale, M. E., Hedrick, T. L., Lappin, A. K., Nichols, T. R., Quinn, R. D., Satterlie, R. A. & Szymik, B. (2007), Neuromechanics: an integrative approach for understanding motor control. *Integrative and Comparative Biology* **47** (1), 16–54.

Niv, Y. (2019), Learning task-state representations. *Nature Neuroscience* **22** (10), 1544–1553.

Niven, J. E., Ott, S. R. & Rogers, S. M. (2012), Visually targeted reaching in horse-head grasshoppers. *Proceedings of the Royal Society B: Biological Sciences* **279** (1743), 3697–3705.

Norman, D. A. & Shallice, T. (1986), Attention to Action: Willed and Automatic Control of Behavior. In *Consciousness and Self-Regulation* (ed. R. J. Davidson, G. E. Schwartz & D. Shapiro), pp. 1–18. Springer US, Boston, MA.

Nuin, Y. L. E., Lopez, N. G., Moral, E. B., Juan, L. U. S., Rueda, A. S., Vilches, V. M. & Kojcev, R. (2019), ROS2Learn: a reinforcement learning framework for ROS 2.

Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. (2015), Hippocampal place cells construct reward related sequences through unexplored space. *eLife* **4**, e06063.

Orlovsky, G., Deliagina, T. & Grillner, S. (1999), *Neuronal Control of Locomotion: From Mollusc to Man*. Oxford University Press.

Page, K. L., Zakotnik, J., Dürr, V. & Matheson, T. (2008), Motor Control of Aimed Limb Movements in an Insect. *Journal of Neurophysiology* **99** (2), 484–499.

Paskarbeit, J., Schilling, M., Schmitz, J. & Schneider, A. (2015), Obstacle crossing of a real, compliant robot based on local evasion movements and averaging of stance heights using singular value decomposition. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3140–3145.

Pearson, K. G. (1995), Proprioceptive regulation of locomotion. *Current Opinion in Neurobiology* **5** (6), 786–791.

Peng, X. B., Chang, M., Zhang, G., Abbeel, P. & Levine, S. (2019), Mcp: Learning composable hierarchical control with multiplicative compositional policies. In *Advances in Neural Information Processing Systems 32* (ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox & R. Garnett), pp. 3686–3697. Curran Associates, Inc.

Pulvermüller, F. (2018), Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology* **160**, 1–44.

Ritzmann, R. E. & Büschges, A. (2007), Adaptive motor behavior in insects. *Current Opinion in Neurobiology* **17** (6), 629 – 636.

Rizzolatti, G. (2005), The mirror neuron system and its function in humans. *Anatomy and Embryology* **210** (5–6), 419–421.

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996), Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* **3** (2), 131–141.

Romano, D., Marini, F. & Maravita, A. (2017), Standard body-space relationships: Fingers hold spatial information. *Cognition* **165**, 105–112.

Schaal, S. & Schweighofer, N. (2005), Computational motor control in humans and robots. *Current Opinion in Neurobiology* **15** (6), 675–682.

Schilling, M. (2011*a*), Learning by seeing—associative learning of visual features through mental simulation of observed action. In *Proc. of the ECAL 2011, Paris* (ed. R. Doursat), pp. 731–738. MIT Press.

Schilling, M. (2011*b*), Universally manipulable body models — dual quaternion representations in layered and dynamic MMCs. *Autonomous Robots* **30** (4), 399–425.

Schilling, M. (2017), Old Actions in Novel Contexts — a Cognitive Architecture for Safe Explorative Action Selection. In *Proceedings of the Artificial Intelligence and Simulation of Behaviour Conference (AISB 2017)*.

Schilling, M. (2019*a*), Hierarchical Dual Quaternion-based Recurrent Neural Network as a flexible internal body model. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.

Schilling, M. (2019*b*), Setup of a recurrent neural network as a body model for solving inverse and forward kinematics as well as dynamics for a redundant manipulator. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.

Schilling, M., Chang, N., Rohlfing, K. J. & Spranger, M. (2020), Simulation across representation: The interplay of schemas and simulation-based inference on different levels of abstraction. *Behavioral and Brain Sciences* **in press**.

Schilling, M. & Cruse, H. (2012), What's next: Recruitment of a grounded predictive body model for planning a robot's actions. *Frontiers in Psychology* **3** (383).

Schilling, M. & Cruse, H. (2016), Avoid the hard problem: Employment of mental simulation for prediction is already a crucial step. *Proceedings of the National Academy of Sciences* **113** (27), E3811–E3811.

Schilling, M. & Cruse, H. (2017), Reacog, a minimal cognitive controller based on recruitment of reactive systems. *Frontiers in Neurorobotics* **11**.

Schilling, M. & Cruse, H. (2020), Decentralized control of insect walking - a simple neural network explains a wide range of behavioral and neurophysiological results. *PLoS Computational Biology* **16** (4), e1007804.

Schilling, M., Cruse, H. & Arena, P. (2007), Hexapod Walking: an expansion to Walknet dealing with leg amputations and force oscillations. *Biological Cybernetics* **96** (3), 323–340.

Schilling, M., Hoinville, T., Schmitz, J. & Cruse, H. (2013*a*), Walknet, a bio-inspired controller for hexapod walking. *Biological Cybernetics* **107** (4), 397–419.

Schilling, M., Konen, K. & Korthals, T. (submitted, 2020*a*), Decentralized deep reinforcement learning for a distributed and adaptive locomotion controller of a hexapod robot.

Schilling, M. & Melnik, A. (2018), An Approach to Hierarchical Deep Reinforcement Learning for a Decentralized Walking Control Architecture. In *Biologically Inspired Cognitive Architectures 2018. Proceedings of the Ninth Annual Meeting of the BICA Society, 848*.

Schilling, M. & Narayanan, S. (2013), Communicating with executable action representations. In *2013 AAAI Spring Symposium Series*.

Schilling, M., Paskarbeit, J., Hoinville, T., Hüffmeier, A., Schneider, A., Schmitz, J. & Cruse, H. (2013*b*), A hexapod walker using a heterarchical architecture for action selection. *Frontiers in Computational Neuroscience* **7**, 126.

Schilling, M., Paskarbeit, J., Ritter, H., Schneider, A. & Cruse, H. (submitted, 2020*b*), From adaptive locomotion to predictive action – cognitive control for a six-legged walker.

Schilling, M., Paskarbeit, J., Schmitz, J., Schneider, A. & Cruse, H. (2012), Grounding an Internal Body Model of a Hexapod Walker — Control of Curve Walking in a Biological Inspired Robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2762–2768.

Schilling, M., Paskarbeit, J., Schneider, A. & Cruse, H. (2012), Flexible internal body models for motor control: On the convergence of constrained dual quaternion mean of multiple computation networks. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.

Schilling, M., Ritter, H. & Ohl, F. W. (2019), From Crystallized Adaptivity to Fluid Adaptivity in Deep Reinforcement Learning — Insights from Biological Systems on Adaptive Flexibility. In *IEEE International Conference on Systems, Man, and Cybernetics 2019*.

Schmitz, J., Schneider, A., Schilling, M. & Cruse, H. (2008), No need for a body model: Positive velocity feedback for the control of an 18-DOF robot walker. *Applied Bionics and Biomechanics, Special Issue on Biologically Inspired Robots* **5** (3), 135–147.

Schneider, W., Albert, J. & Ritter, H. (2020), Enabling cognitive behavior of humans, animals, and machines: A situation model framework. *ZiF-Mitteilungen* **1**, 21–34.

Schneider, W. & Chein, J. M. (2003), Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science* **27** (3), 525–559.

Schneider, W. & Shiffrin, R. M. (1977), Controlled and automatic human information processing. *I. Detection, search, and attention. Psychological Review* pp. 1–66.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. (2017), Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* .

Shadmehr, R. & Wise, S. P. (2005), *The Computational Neurobiology of Reaching and Pointing : A Foundation for Motor Learning (Computational Neuroscience)*. The MIT Press.

Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H. & Jaffar, A. (2011), Humanoid robot NAO: Review of control and motion exploration. In *2011 IEEE International Conference on Control System, Computing and Engineering*, pp. 511–516.

Shenoy, K. V., Sahani, M. & Churchland, M. M. (2013), Cortical Control of Arm Movements: A Dynamical Systems Perspective. *Annual Review of Neuroscience* **36** (1), 337–359.

Shiffrin, R. M. & Schneider, W. (1977), Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theoryerceptual learning, automatic attending and a general theory. *Psychological Review* **84** (2), 127–190.

Sponberg, S. & Full, R. J. (2008), Neuromechanical response of musculo-skeletal structures in cockroaches during rapid running on rough terrain. *The Journal of Experimental Biology* **211** (Pt 3), 433–446.

Stanovich, K. E. & West, R. F. (2000), Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* **23**, 645–665.

Steels, L. (2003), Intelligence with representation. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **361** (1811), 2381–2395.

Storrs, K. R. & Kriegeskorte, N. (2019), Deep learning for cognitive neuroscience. *arXiv preprint arXiv:1903.01458* .

Strauss, R. & Pichler, J. (1998), Persistence of orientation toward a temporarily invisible landmark in Drosophila melanogaster. *Journal of Comparative Physiology A* **182**, 411–423.

Sutton, R. S. & Barto, A. G. (2018), *Reinforcement Learning: An Introduction*, 2nd edn. The MIT Press.

Sutton, R. S., Precup, D. & Singh, S. (1999), Between MDPs and semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning. *Artificial Intelligence* **112** (1-2), 181–211.

Theunissen, L. M., Vikram, S. & Dürr, V. (2014), Spatial co-ordination of foot contacts in unrestrained climbing insects. *Journal of Experimental Biology* **217** (18), 3242–3253.

Tingley, D. & Buzsáki, G. (2018), Transformation of a spatial map across the hippocampal-lateral septal circuit. *Neuron* **98**, 1229–1242.E5.

Todorov, E. & Jordan, M. I. (2002), Optimal feedback control as a theory of motor coordination. *Nature Neuroscience* **5** (11), 1226–1235.

Tuthill, J. C. & Wilson, R. I. (2016), Mechanosensation and Adaptive Motor Control in Insects. *Curr Biol* **26** (20), R1022–R1038.

Uithol, S., van Rooij, I., Bekkering, H. & Haselager, P. (2012), Hierarchies in Action and Motor Control. *Journal of Cognitive Neuroscience* **24** (5), 1077–1086.

Ullman, S. (2019), Using neuroscience to develop artificial intelligence. *Science* **363** (6428), 692–693.

Vezhnevets, A. S., Osindero, S., Schaul, T., Heess, N., Jaderberg, M., Silver, D. & Kavukcuoglu, K. (2017), Feudal networks for hierarchical reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, pp. 3540–3549.

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., Binder, J. R., Men, W., Gao, J.-H. & Bi, Y. (2018), Organizational Principles of Abstract Words in the Human Brain. *Cerebral Cortex* **28** (12), 4305–4318.

Webb, B. (2004), Neural mechanisms for prediction: do insects have forward models? *Trends in Neurosciences* **27** (5), 278–282.

Webb, B. (2020), Robots with insect brains. *Science* **368** (6488), 244–245.

Wessnitzer, J. & Webb, B. (2006), Multimodal sensory integration in insects—towards insect brain control architectures. *Bioinspiration \& Biomimetics* **1** (3), 63–75.

Whelan, P. J. (1996), Control of locomotion in the decerebrate cat. *Progress in neurobiology* **49** (5), 481–515.

Wilson, M. (2008), How did we get from there to here? An evolutionary perspective on embodied cognition. In *Handbook of Cognitive Science: An Embodied Approach* (ed. P. Calvo & A. Gomila), pp. 375–393. Academic Press.

Wolpert, D. M. & Flanagan, J. (2001), Motor prediction. *Current Biology* **11** (18), 729–732.

Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995), An internal model for sensorimotor integration. *Science* **269**, 1880–1882.

Wolpert, D. M. & Kawato, M. (1998), Multiple paired forward and inverse models for motor control. *Neural Networks* **11** (7-8), 1317–1329.

Wosnitza, A., Bockemühl, T., Dübbert, M., Scholz, H. & Büschges, A. (2013), Inter-leg coordination in the control of walking speed in drosophila. *Journal of experimental biology* **216** (3), 480–491.

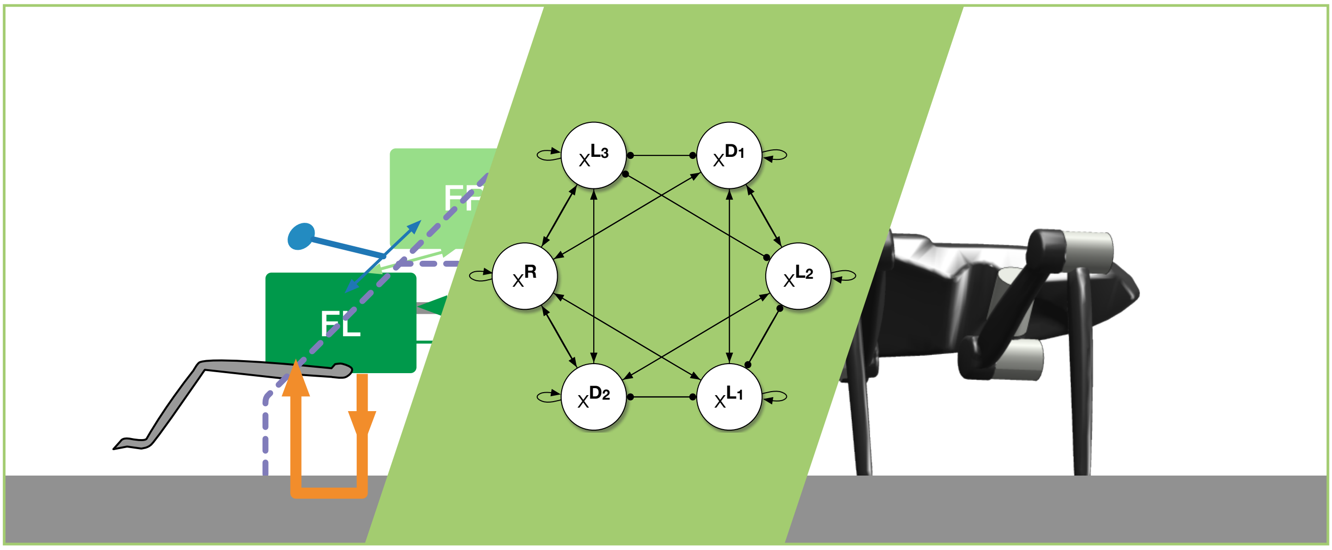
Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. (2017), Hippocampal awake replay in fear memory retrieval. *Nat Neurosci* **20** (4), 571–580.

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. L. & Wood, R. (2018), The grand challenges of Science Robotics. *Science Robotics* **3** (14), eaar7650.

Yoo, Y. & Kim, W. (2017), On Decoding Grid Cell Population Codes Using Approximate Belief Propagation. *Neural Computation* **29** (3), 716–734.

Zhou, Y., Barnes, C., Lu, J., Yang, J. & Li, H. (2019), On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5745–5753.

Zill, S. N. & Moran, D. T. (1981), The Exoskeleton and Insect Proprioception: III. Activity of Tribal Campaniform Sensilla During Walking in the American Cockroach, Periplaneta Americana. *Journal of Experimental Biology* **94** (1), 57–75.



Decentralization and Hierarchical Organization for Control of Adaptive and Cognitive Behavior in Autonomous Robots

Cognition—understood as a form of planning ahead—complements adaptive behavior. It leverages knowledge about performing a specific behavior into a novel context while minimizing any harm to the behaving system itself as it is using an internal simulation to predict possible outcomes. In this thesis, I propose a minimal cognitive system that integrates these two kinds of processes in one control system for a six-legged robot.

On the one hand, adaptive behavior emerges from interaction of simple local control modules which allows the system to react quickly when facing disturbances. Detailed experimental findings in insects suggests that this evolved flexibility results from a hierarchical and decentralized architecture. While a lower control level coordinates muscle activation patterns and joint movements on a short timescale, a higher level handles action selection on longer timescales.

On the other hand, following a bottom-up approach this is extended towards a cognitive system that is able to invent new behaviors and to plan ahead. Using a grounded internal body model planning is realized as a form of internal simulation of possible actions which are applied out of their original context. Exploiting the decentralized architecture, this cognitive expansion allows to test and predict properties of newly invented behaviors, while the body is decoupled from the control system.

The thesis introduces the minimal cognitive system as it is applied on the robot Hector in a climbing task. It consecutively introduces the underlying control characteristics and relates these to findings from biology and neuroscience. First, hierarchical organization can be found in many animals and it structures control into parsimonious modules. Second, this is complemented by research on stick insects in particular which offers an even more detailed neuronal and behavioral level for analysis. This emphasizes decentralization of control structures and the importance of an embodied perspective which integrates bodily properties into the concurrent control process exploiting, for example, elasticities of muscles for simplifying the control problem. Third, internal representations are introduced in a bottom-up manner as grounded internal models—realized as recurrent neural networks—that are at first considered in the context of serving a specific behavior. Fourth, as a consequence, cognitive processing is realized as recruitment of the already existing flexible internal models in an internal simulation. The underlying architecture is applied on the hexapod robot Hector and analyzed in detail in simulation. Furthermore, learning is considered for this approach.