Contours of Cognition

Tjøstheim, Trond A.

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Tjøstheim, T. A. (2022). *Contours of Cognition* (1 ed.). [Doctoral Thesis (compilation), Department of Philosophy]. Lund University.

Total number of authors:
1

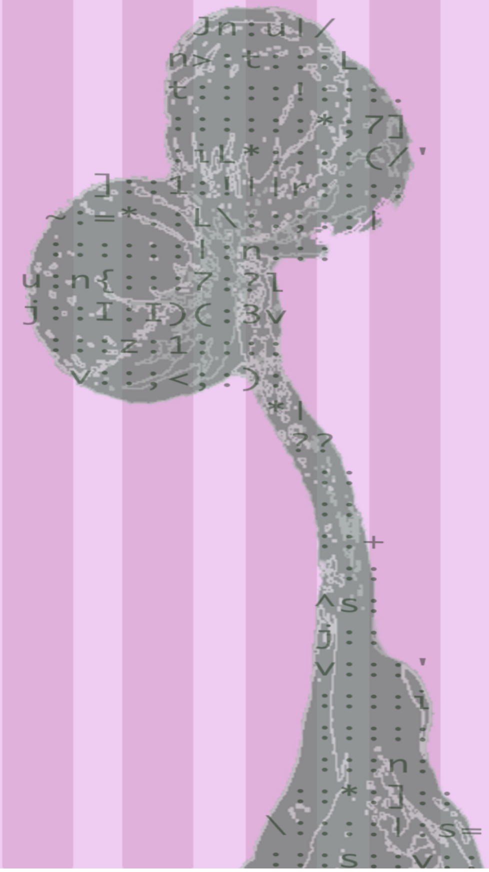
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Contours of Cognition

TROND A. TJØSTHEIM

COGNITIVE SCIENCE | DEPARTMENT OF PHILOSOPHY | LUND UNIVERSITY



Contours of Cognition

Contours of Cognition

by Trond A. Tjøstheim

LUND
UNIVERSITY

Doctoral Dissertation
Thesis advisors: Prof. Christian Balkenius, Dr. Birger Johansson
Faculty opponent: Prof. Martin V. Butz

To be presented, with the permission of the Faculty of Humanities and Theology of Lund University, for public criticism at LUX (C121, Ground floor, LUX building) on Monday, the 19th of September 2022 at 10:15.

DOKUMENTDATABLAD enl SIS 61 41 21

Organization
LUND UNIVERSITY

Department of Philosophy
Box 192
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2022-09-19

Sponsoring organization

Author(s)
Trond A. Tjøstheim

Title and subtitle
Contours of Cognition

Abstract

This thesis concerns the nature of cognition. It posits that cognitive processes primarily are means to maintain allostasis in organisms whose ecological niches require movement to approach food-resources and avoid predation. Hence triggering, or motivation, of behaviours are a consequence of prediction errors from the body resulting from biological variables moving away from homeostasis.

Depending on circumstance and the nature of the particulars of the ecological niche, an organism may require the ability to find the way to a goal-site containing food or water, perceive its surroundings in order to trigger allostatic behaviour, make choices and priorities, and predict outcomes. Hence, cognition is situated in a larger context of staying alive, but efforts are also made to zoom in on exactly how some important cognitive processes may plausibly work, on the level of neural units and networks. These processes include visual perception, spatial cognition, predictive simulation processes (intelligence), and familiarity based trust, as well as reflection, decision-making, and memory.

Key words

computational cognitive neuroscience, perception, memory, spatial cognition, trust, pupil dynamics, reflection, thinking, intelligence, reinforcement learning, foraging, allostasis, decision making

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title 1101-8453 Lund University Cognitive Studies 183

ISBN
978-91-89415-23-2 (print)
978-91-89415-24-9 (pdf)

Recipient's notes

Number of pages
282

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2022-05-25

Contours of Cognition

by Trond A. Tjøstheim

LUND
UNIVERSITY

Cover illustration front: Mycelium fruiting body by Trond A. Tjøstheim.

Cover illustration back: Multiresolution receptive fields by Trond A. Tjøstheim.

Faculty of Humanities and Theology
Department of Philosophy
Cognitive Science

ISBN: 978-91-89415-23-2 (print)
ISBN: 978-91-89415-24-9 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2022

Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

To my family, and
to hope:
that spring of springs.

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **From Focused Thought To Reveries: a Memory System For A Conscious Robot**
C. Balkenius, **T. A. Tjøstheim**, B. Johansson, P. Gärdenfors
Frontiers in Robotics and AI, 2018, pp. 1-29

II **Cumulative Inhibition In Neural Networks**

T. A. Tjøstheim, C. Balkenius
Cognitive Processing, 2019, 20(1), pp. 87-102

III **A computational model of trust-, pupil-, and motivation dynamics**

T. A. Tjøstheim, B. Johansson, C. Balkenius
Proceedings of the 7th International Conference on Human-Agent Interaction, 2019, pp. 79-185

IV **The Cognitive Philosophy of Reflection**

A. Stephens, **T. A. Tjøstheim**
Erkenntnis, 2020, pp. 1-24

V **The Missing Link Between Memory and Reinforcement Learning**

C. Balkenius, **T. A. Tjøstheim**, B. Johansson, A. Wallin, P. Gärdenfors
Frontiers in Psychology, 2020, 11, p. 3446

VI **Intelligence as Accurate Prediction**

T. A. Tjøstheim, A. Stephens
Review of Philosophy and Psychology, 2021, pp. 1-21

VII **Direct Approach or Detour: A Comparative Model of Inhibition and Neural Ensemble Size in Behavior Selection**
T. A. Tjøstheim, B. Johansson, C. Balkenius
Frontiers in Systems Neuroscience, 2021, 15

All papers are reproduced with permission of their respective publishers.

Preface

Nature is not limited by the boundaries of human categories. Hence what may be interpreted by humans as ”cognitive” may appear at first clear cut, but becomes less so when zooming into the details. In this thesis I want to arrive at an understanding of what cognition is, and what it may constitute. The result is by necessity a coarse approximation of what is going on in organisms. I build on the literature of cognitive science, though this thesis reflects my own attempt to trace a thread from the inside of neurons, to the behaviour of organisms, via the signal flow of networks. Still, it is not a fully fledged theory, only an interpretation, and a suggestion of how things *might* hang together.

The experiments conducted for this thesis rely on computational simulations, and have different levels of resolution. Some use spiking neurons that behave approximately like biological neurons, while others use more abstract computational representations. Although only one experiment uses an agent of sorts, it is virtual, moving in a virtual world. The overall aim of the simulations is nevertheless that they should inform models that can control robots that move in the real world and interact with humans.

Computational models have limitations as scientific tools since unless fitted to empirical data, they risk becoming arbitrary and lose predictive power. The wonderful thing about computational models is that they allow exploration and the trying out of ideas. Compared to theoretical synthesis, they encourage also concretization, since the model must after all compile and run on the computer. At their best, computational models suggest how things could possibly work; fitted to data they might also propose how things could *reasonably* work. Alas, asking if they can answer how things *actually* are, is usually demanding too much.

Nevertheless, I hope that the perspectives presented herein may be of some use, and provide inspiration to explore.

Acknowledgements

First I must thank my family, without whom this project would be impossible.

Then I want to thank my supervisors, Christian Balkenius and Birger Johansson for their patience, help, and friendship during this time.

I have worked and discussed with many, if not most of the staff, students, and researchers at LUCS, past and present; warm thanks to Erik Anderberg, Katarzyna Bobrowicz, Valentina Fantasia, Zahra Gharaee, Kerstin Gidlöf, Agneta Gulz, Peter Gärdenfors, Magnus Haake, Lars Hall, Jana Holsanova, Ivo Jacobs, Petter Johansson, Can Kabadayi, Amandus Krantz, Maybí Morell Ruiz, Manuel Oliva, Kristin Osk Ingvarsdottir, Helena Osvath, Mathias Osvath, Tomas Persson, Stephan Reber, Thomas Rejsenhus Jensen, Thomas Strandberg, Björn Sjödén, Eva Sjöstrand, Alexander Tagesson, Eva–Maria Ternblad, Thibault Boehly, Matthew Tompkins, Annika Wallin, and Claudia Zeiträg.

A big thanks to all the administrative staff, tirelessly helpful for a confused student: Tobias Hansson Wahlberg, Martin Jönsson, Ingela Byström, Anna Cagnan Enhörning, Fredrik Eriksson, Annah Smedberg-Eivers, and Anna Östberg.

Thank you to everyone at the department of philosophy for interesting discussions! In particular I want to thank Erik Olsson, Robin Stenwall, Max Minden Ribeiro, Frits Gåvertsson, Matthias Gunnemyr, Gloria Mähringer, Signe Savén, Marta Johansson Werkmäster and Jakob Werkmäster, Justine Jacot and Emmanuel Genot, Henrik Andersson, Melina Tsapos, Martin Sjöberg, Alexander Velichkov, Jiwon Kim, Jenny Magnusson, Robert Pál-Wallin, Anton Emilsson, and Niklas Dahl. Also thanks to Arthur Schwaninger at the University of Zürich for cooperation and discussions.

Some have become particularly close during these years, and I am very grateful for the time with Andreas Stephens, Andrey Anikin, Kalle Palm, Jakob Stenseke, Gabriel Vogel, Maximilian Roszko, and Rasmus Arnling Bååth.

Similarly I want to acknowledge all my friends in Stavanger and elsewhere for your love and support. I hope I will see you more often soon!

I must give special thanks to my friends with which I share an office, and who have become precious to me: Samantha Stedtler, Jens Nirme, and Betty Tärning.

Finally, I want to acknowledge and thank Elizabeth Haugland at *Høgskolen i Bergen* for giving me the idea to at all attempt a PhD.

Populärvetenskaplig sammanfattning på svenska

Denna avhandling handlar om hur kognition fungerar, och vad dess syfte är. Avhandlingens ramverk bygger på att kognitiva processer i första hand medverkar till att hålla organismer vid liv, och detta särskilt i miljöer som kräver att organismen förflyttar sig. Det innebär att en organism inte kan sitta still och få maten serverad. Istället måste organismen bege sig ut i världen och aktivt hitta något att äta. Detta innebär att ett beteende sätts igång när kroppens fysiologiska variabler avviker från de värden som håller organismen vid god hälsa, det vill säga, att kroppen inte längre är i *homeostas*.

Beroende på situationen, och den ekologiska nischens karaktär, behöver en organism hitta en väg som leder till något att äta och en väg som leder till något att dricka. För att lyckas med detta, måste organismen kunna tolka sin omvärld, kunna göra olika val, prioritera, samt förutse vad konsekvenserna kommer att blir av de val den gör. Härigenom sätts kognition i kontexten av att hålla organismen vid liv. Avhandlingen går även i detalj in på hur kognitiva processer fungerar på nätverks- och cellnivå. Sådana processer innefattar visuell perception, spatial kognition, prediktiva simuleringsprocesser, intelligens, och tillit baserad på familjäritet. Även tänkande, i form av reflektion, beslutsfattande, och minnesprocesser behandlas i avhandlingen. Slutligen diskuteras hur artificiella kognitiva system sannolikt kan skapas, och att essensen av kognition är förmågan att känna igen, men även förmågan att kunna välja.

Contours of Cognition

Chapter 1

Introduction

This thesis is about how cognition may be considered a tool for maintaining allostatic processes, specifically for species that have evolved to live in ecological niches that require movement in space. That is, cognition makes foraging possible. This is in contrast with organisms which do not forage, like sponges, or sea–annelids, which rely on nutrients drifting in water.

Consider now the slime mould. This organism, which can be found in many forest environments, feeds on bacteria and on plant material Vallverd'u et al. (2018). When food is abundant many slime moulds live as single celled organisms, but can aggregate together when food becomes scarce, and produce fruiting bodies which release spores. What makes slime mould interesting in the context of cognition, is that these organisms have been shown to do apparently remarkable feats of advanced computation. In particular, they are very good at finding the shortest path between points. Hence they are beloved of many mathematicians working with graph theory, as well as engineers planning rail way systems. Can these feats be counted as cognition? Intuitively this categorization seems obviously false. Slime moulds lack even a rudimentary nervous system, let alone a brain. But how then can they outperform many computers, and certainly human beings when solving complex network problems?

In particular, the slime mould is an example of how *oscillation*s and *rhythms* play parts in cognition. This is important, since neurons are also oscillatory and rhythmic entities. Neurons are not on or off in the same way that digital electronics are on or off; they produce trains of *action potentials* that can be emitted at high or low rates; neurons never become entirely silent, unless they are damaged. Neurons are always found in groups, and cannot exist as singletons. Even the smallest of nerve systems, e.g. that of the microscopic aqueous rotifer, consist of several tens of neurons

(Kotikova, Raikova, Reuter, & Gustafsson, 2005). This means that the dynamics of a nervous system have similarities with a musical band, where rhythms of loud players can impose on more quiet ones. Like musicians in a musical band, some neuronal populations play the base rhythms, while others play faster melody–like rhythms on top of the basic beat. Unlike a typical band though, neuronal populations can be inhibited by each other, and be affected by *neuromodulators*. These neuromodulators can determine whether individual neurons are more or less active, and the rate at which their synapses adapt.

Both slime moulds, and organisms with nervous systems, employ oscillation to maintain their life processes. Indeed, oscillators are primary elements of any living system; even if an organism appears to be sitting still, it is teeming inside with waves and oscillations. A fact of physiology is also that organisms are dependent on limits. Life processes can only keep going in environments that are *regulated*: neither too cold nor too hot, neither too alkaline nor too acidic. Regulation is thus one of life's hallmarks. Both in psychology and in cognitive science, this may be under–appreciated. In particular, the computer metaphor with its focus on information processing abstract from biology, makes it easy to miss the fundamental regulatory requirements of living systems. The Cartesian paradigm famously summarized as *"I think therefore I am"* is in this context not helpful: Rather than living to process information, organism process information to stay alive. As we will see below, organisms tend also to process as little information as they can get away with, since information processing is an expensive endeavour. Although the ideas regarding how regulation relates to cognition may not be as widely spread as one might hope, they are not at all new. In the next section, I present a very abridged history.

1.1 A brief history of cybernetics and physiological control

Cybernetics may be considered the study of feedback control as applied to biological systems. This section concerns what might be seen as the "cybernetic theory of cognition", or the notion that cognitive processes partake in the physiological regulation of the body, and that these processes are themselves subject to regulation.

Although the term "cybernetics" originated with Wiener (1948), the notion that biological bodies are regulated can at least be traced to Cannon (1932). In this book, Cannon introduces the term *homeostasis* to denote physiological equilibrium states in the body. The essence of homeostasis is the notion that organisms are dependent on their internal environments being neither too little or too much of anything. For example, both body temperature and sodium solution in the blood have homeostatic values. If these values drift away, various regulatory processes will be engaged to counter the drift.

Prior to Cannon, Whitehead (1929) proposed that the fundamental aspect of reality is *change*, or processes, in contrast to static notions of unchangeable building blocks. This is in turn similar to the ancient Greek concept of *flux*, some times attributed to Heraclitus (Barnes, 1982).

von Uexküll (1934) proposed that animals inhabit an *Umwelt*, conditioned on their sensory apparatus. This means that organism cannot sense the totality of their environment, only that small part of it that makes a difference to their survival and procreation. According to von Uexküll, the ”meaning” of signals is thus the way they affect physiology and behaviour.

Craik (1943) proposed the notion of internal mental models, and the comparative ease of handling relative magnitudes compared to absolutes. In his proposal, thought was considered akin to the workings of a machine, able to model or mirror external goings on.

Wiener (1948) used concepts from engineering and mathematics having to do with stabilization and regulation by means of signal feedback, to propose that the nervous system too makes use of feedback flows to stabilize itself.

Building on Wiener’s work, Ashby (1952) formulated the problem of how the brain can mediate adaptive behaviour. Focusing on learned behaviour in contrast to reflexive behaviour, Ashby sought to explain how feedback mechanisms can contribute to adaptation. He also pointed attention to the contrast between equilibrium seeking which will gradually reduce the number of states a system can occupy, and how injecting energy into the system by amplification, can increase the number of available states.

Crosson and Sayre (1968) and Sayre (1976) engaged specifically with communication theory Shannon (1948) and information theory Brillouin (1956) in the context of cybernetics. They concerned themselves in particular about philosophy of mind, and aspects of human thinking, like what constitutes meaning, memory, and agency. In the same time period, Lorenz (1977) followed Cannon, von Uexküll, Wiener, and others to theorize about the evolution of mind in relation to survival and homeostasis. Being a biologist, he connected these concepts also directly to experiments with animals.

Friston, Kilner, and Harrison (2006) takes up Craik’s notion of an internal model, and proposes that the brain specifically produces predictions, which when failing will produce prediction errors. According to the authors, these *prediction errors* can be formulated by means of *informational free energy*. In the context of cybernetics, prediction errors may be understood as relative to the system's reference values, or setpoints. The formulation of free energy proposed by Friston, uses the mathematical equivalence of difference between two probability distributions used in thermodynamics to find how much energy is free to do work when heat energy is subtracted. In information theory, free energy is in turn used to denote information that is not shared between two systems. The commonality between the thermodynamic and information theoretic notions of free energy was also noted by Sayre (1976), though less explicitly. In the author's theory, the informational difference is realized as activity in neurons that can be used to drive adaption through learning. Notably, this work transcends philosophy and proposes mathematical models for the proposed relationships, and also anchors the theory in brain science.

Liu, Slotine, and Barabási (2011) updates control theory for network applications, showing how feedback control can be generalized by means of *driver nodes*. In a network, these nodes act as effectors, and with a suitable control signal and sufficient energy, can push the rest of the network into specific states that would otherwise not be easily reachable.

Shi Gu et al. (2015) applied the theory presented by Liu et al. (2011) in an experiment on humans. Based on brain imaging, they found that brain networks display *controllability*, or possibility to have their activity manipulated as part of driving the whole system along a trajectory. The results of the experiment and of the analyses performed by the authors, indicate that areas that are densely connected, particularly in the *default mode network* contribute to moving the system into states that have *low* energy demands. Weakly connected areas on the other hand, contribute to moving the system into states that have *high* energy demands, or that are more difficult to reach. Making such transitions implies expending effort to energize driver nodes. The authors also found that areas related to attention, which are located on the boundary between network communities contribute to integrate or segregate neural populations. Related to cognitive control, they found that brain areas with high average controllability require less cognitive effort to manipulate.

Butz (2016) integrates several theoretical frameworks of cognition and computation, including those of informational free energy (Friston et al., 2006), predictive coding (Rao & Ballard, 1998), anticipatory behaviour (Hoffmann, 2003), event segmentation (Hommel, Müsseler, Aschersleben, & Prinz, 2001), and cognitive development (Konczak, Borutta, Topka, & Dichgans, 2004), with motivation based on homeostasis. This author then uses a computational, and machine learning perspective, to propose how cognition may be understood. Much of the work presented here reiterates this integrative proposal, but with less emphasis on computation and in favor of neuroscience.

1.2 Overview of thesis

This thesis is organized as follows. In chapter 2, I explore what might be the fundamental processes of cognition, using as a starting point the proposal that recognition and choice may be at the center. This chapters covers ground from non–biological processes involving thermodynamics and oscillation, to cognitive processes like perception, recognition, and choice.

In chapter 3, I go into physiological fundamentals of life maintaining processes, the target states of those processes, and the fuel that drives them. In the language of physiology, these are called respectively allostasis, homeostasis, and chemical energy. The reason for this focus on physiology is that they are the origin of all higher level processes, including cognition. Understanding this origin allows better understanding also of what limits, and drives, both mental– and motor behaviour. I relate these physiological processes to cognition at the organismic level by arguing that behaviour may be considered allostatic, and at the level of the central nervous system, by going into three particular neuromodulator systems that are important for foraging behaviour.

In chapter 4, I continue on the theme of allostasis and connect it more explicitly to regulation. I show how the cybernetics framework and control theory are useful perspectives on allostasis, and explain the importance of feedback in biological systems. Allostasis is connected to cognition in this chapter via hunger as a motivator for foraging.

After this, in chapter 5, I focus on the roles that selection, attention, and choice plays in behaving organisms. At this point we are firmly in the arena of cognition, and I go into some detail about how attention may be viewed as a means to keep cognitive processes tractable for individual organisms. Finally in this chapter, I relate choice processes to both behaviour and goals, and explain how choice relies on value.

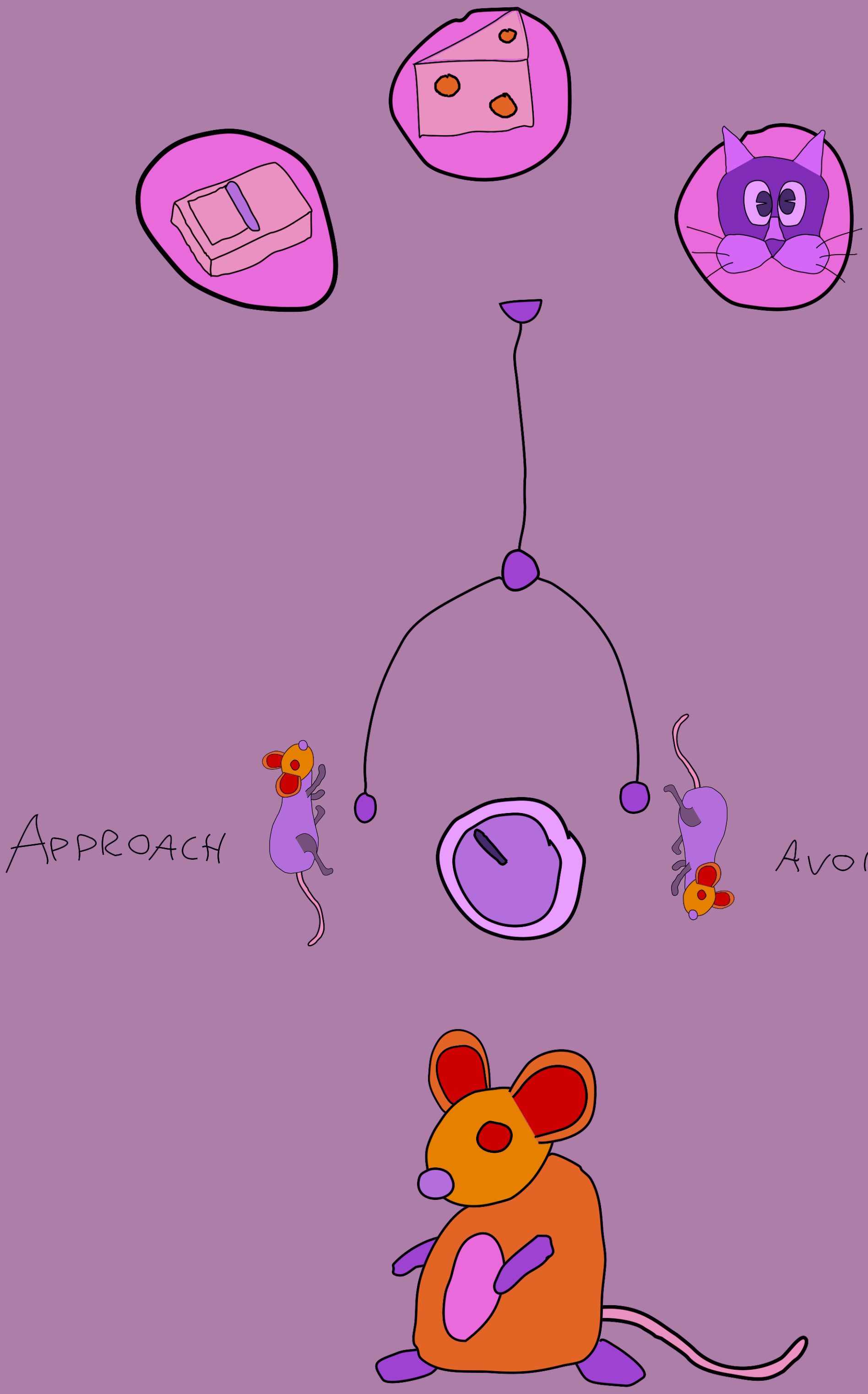
When a choice has been made about where to go, the next task is to navigate to this chosen place. In chapter 6, I revisit how allostasis drive foraging, and explore the role of learning to support navigation. At the end of this chapter, I also suggest that machinery for navigating to foraging places could be appropriated to serve more general purposes of predicting all kinds of situational–dependent dynamics.

Nature commonly places obstacles in the way of foraging animals. Being able to move, bend, reach, and grasp can thus be helpful. In chapter 7, I look into how the variability and complexity of environments may drive aspects of cognition, such as abilities to generalize and predict. Then I connect these abilities to manipulation behaviour in particular. The ability to grasp opens up for tool use. Tool use may again increase the need for planning, since tools may need to be carried to where they are needed.

Finally in this chapter, I suggest how abstract problem solving may be mediated by foraging related manipulation abilities.

In chapter 8, I make the transition from biology to technology and explore what an artificial cognitive system might look like, and how to interpret allostasis in such a system.

In the final chapter 9, I attempt to draw together the strands from the previous chapter to form contours of cognition. I discuss how cognition might be understood as existing on a scale, rather than being a categorical ability. Finally, I discuss the issue of reflection and thinking, and how processes like these might be useful for integrating various sources of knowledge, and contribute to making better choices in future.


APPROACH
AVOID

Chapter 2

Fundamental processes for cognition

Cohen (1992), Cohen (2000) and Wallace and Wallace (2017) propose that the fundamental building block processes of cognition include *recognition* and *choice*. Here I use this proposal as a starting points in an attempt to get a grip on what cognition is. Cohen (2000) used cognition in the context of the immune system, arguing that the process of differentiating between self and other, and engaging appropriate responses, might be interpreted as a type of cognitive process. Note that Cohen's original formulation was that cognitive systems may be characterized by three properties; the first being the ability to choose; the second that cognitive systems contain within them images of their environment; the third that these images are formed from experience. However, I interpret the second and third properties to amount to recognition, since this is a functional way that internal images may be used to control behaviour. It also allows us to consider organisms that do not necessarily use experience to build images, but that rely on evolution to form the images for them.

This chapter is organized in the following way. First, I introduce the notion of thermodynamical gradients, and how such gradients may be a source for ultimately grounding cognitive processes, even if they at first glance may appear to far removed from the cognitive and behavioural level to contribute much to understanding.

Perception is necessary for recognition, and I next describe a model of how perception may develop and some aspects of how it works by developing statistical building blocks.

Memory is also necessary for recognition, in addition to perception, and the two next sections delve into the connections between the two, as well as a reiteration of how recognition and choice together may be considered elementals of cognition.

Oscillators are important mediators of excitation and inhibition, and neurons are prime examples of cells capable of oscillation. In the next section, I discuss how neurons work to support recognition and choice, and how neural ensembles can form motifs that are necessary for larger scale cognitive functions.

Next, I discuss how neural ensembles can tune themselves to both discriminate and generalize, and how neural signals can become meaningful when they relate to behaviour.

Finally in this chapter, I connect recognition to choice and argue along with Cohen and Wallace that recognition and choice may indeed be considered fundamental for cognition.

2.1 Thermodynamical gradients from a behavioural perspective

Gradients may be understood as change spread out over time and space. In figure 2.1 I give some examples of gradients, including gradients of colour, pressure, and geometry. A thermodynamic gradient is typically one of movement in the form of heat, but also of volume, and matter concentration. Note also that heat can be understood as movement, in the sense that low heat entails slow moving particles, while much heat entails quickly moving particles with high *kinetic energy*. An important aspect of thermodynamical gradients are that they tend to reduce over time; concretely this means that if you turn on your oven in a cold room, there will at first be a steep heat gradient in all directions from the oven. But after a while as the room heats up, the difference will be diminished. Similarly, if you add a drop of colour to a glass of water, the colour will be concentrated to a small place at first, but shortly the whole glass will be equally coloured. Both of these processes are due to the second law of thermodynamics, which states that *entropy* must increase over time. Entropy may here be interpreted as disorder, such that the single drop of colour in the glass of water is an example of an ordered configuration, but the whole glass being the same colour is disordered or *entropic*.

In biological cells, thermodynamic gradients exist everywhere, mostly as concentration gradients of molecules and ions on either side of cell membranes, but also inside and outside of compartments inside cells. As mentioned above, heat can be understood as movement. In biological systems, energy is stored as potential movement in the molecule ATP. Without going into excessive details, the ATP molecule works

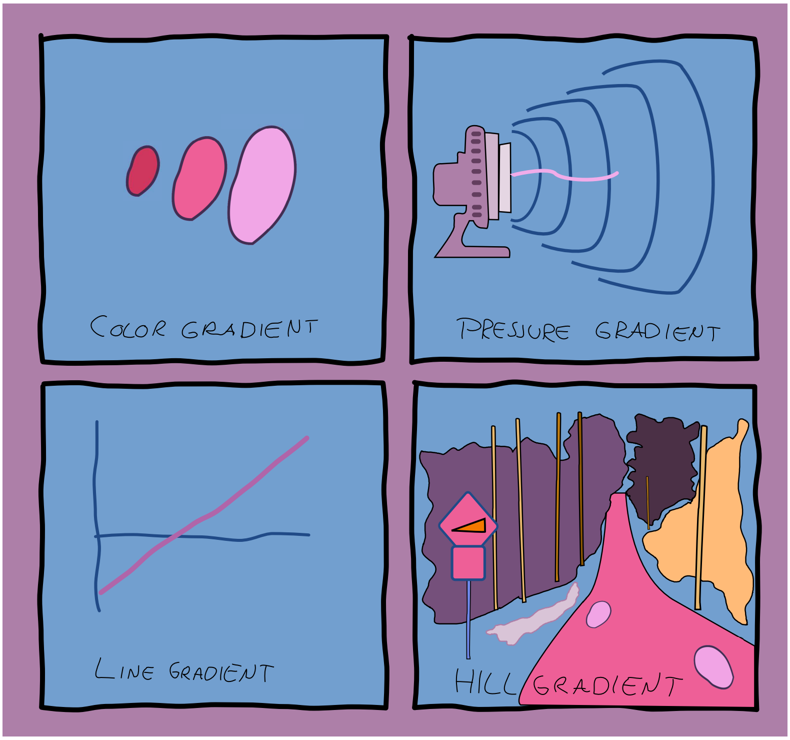


Figure 2.1: Gradients may be understood as change spread out over space or time. **Colour gradients** is the change of colour, here from darker to brighter pink. **Pressure gradients** form when fans operate; the pressure is higher close to the fan than further away from it, thus creating the gradient. **Line gradients** are familiar from geometry, where a line forms an angle with another line. **Hill gradients** naturally form when a landscape has hills and valleys. This example shows a road climbing up a hill.

as a kind of coiled spring that can be triggered when needed to power biochemical processes. But when an ATP molecule is triggered, the resulting movement inevitably affects not only its target but also neighboring structures and molecules. The result is that their movement, or kinetic energy increases. Summed up over the whole cell, we can see that the cell heats up as it uses ATP molecules. The need for using this energy comes from biological processes that have to work in opposition to thermodynamic gradients.

Deacon (2011) refers to processes that go along with thermodynamic gradients as *orthograde*, and those oppose them as *contragrade*. Henceforth I will make use of these short–hands when discussing biological and cognitive processes. In particular, contragrade processes are noteworthy because they imply having to do thermodynamic *work*. This notion of work is similar to what happens after having had a crowded party in your apartment. As your guests move around, they accidentally nudge a picture frame, they drop chips on the floor, they move chairs around. The next day, you have to spend time and effort to return the picture frame and chairs to their original positions, and you have to pick the chips off the floor. This is contragrade work; the nudging and the dropping of chips are orthograde, since they are byproducts of the natural bustle of the party.

In contrast to the orthograde changes of the drop of colour in water, and the air of your heated room, biological systems must continuously do contragrade work to stay alive and avoid having their molecules being spread equally around the environment. What goes on inside a biological cell is a continuous cleaning–up process, making sure that some molecules only stay at certain places and that the internal temperature of the cell is right. As explained above, the energy to power maintenance comes from using ATP. But the energy to make ATP must come from outside the cell, and ultimately from outside the organism. In other words, the organism must eat to stay alive.

If the organism is lucky, food is readily available in its environment, and no particular work need to be done to find it. Some water dwelling organisms, like sponges, get their nutrition from filtering naturally moving water and hence do not need to do much work to stay alive (Keijzer & Arnellos, 2017). Many other organisms, including vertebrates, do need to take action to not die. Hence at the behavioural level, motor behaviour can be triggered by processes that, at the sub–cellular level, include expending work to maintain thermodynamical gradients.

Foraging is a common term used for the behaviour displayed by animals searching for food. Much of this thesis will be dedicated to discussion how cognition may be considered a tool for foraging. At the present time though, I will limit the discussion to some particular examples.

Field mice routinely hunt insects, but are themselves hunted by several predator species (Reznikova, Levenets, Panteleeva, & Ryabko, 2017). This means that a foraging mouse cannot only let itself be preoccupied with finding something to eat, it must also keep itself safe. When foraging, the mouse must find the way to its hunting grounds, identify prey, and catch it. Having eaten, it must scurry back to the safety of its nest. All of these behaviours entail cognition; the mouse must perceive its environment with olfaction, audition, and vision; it must recognize an eatable insect, and predict its behaviour sufficiently to intercept and kill it. Cognitive processes are themselves metabolically costly, but some are more expensive than others. Hence we can expect such processes to be regulated, depending on the what can be gained from employing them. That is, cognitive processes should be subject to *motivation*. In this way, cognition may also be viewed as behaviour that, similar to motor behaviour, are engaged to maintain the thermodynamic gradients existing in all the cells and tissues of an organism.

According to Deacon (2011), cognitive processes may also be consider orthograde or contragrade. An orthograde cognitive process can be exemplified by direct perception, like identifying a path on the forest floor, or finding the entrance to a nest. Such processes are metabolically inexpensive. Contragrade cognitive processes, on the other hand, do require investing energy, and are typically experienced as subjectively effortful. Such processes often involve engaging inhibition, attenuating the flow of action potentials, or redirecting them where they would not otherwise flow. Examples of contragrade behavioural processes for the mouse might be refraining from pouncing on an insect, or walking into an open space where it will be vulnerable to predation. In fact, being able at all to engage in contragrade processes, particularly cognitive processes, tend to be an indication of sophisticated cognition. Typically, employing contragrade cognition is greatly facilitated by motivation (Botvinick & Braver, 2015).

To sum up, this section has shown how thermodynamic gradients, like heat differences, and differences in ion concentration, play a part in higher level processes, including cognition. In contrast to orthograde processes that follow gradients, contragrade processes requires spending energy, e.g. to move ions against a concentration gradient, out of the cell. Animals that cannot rely on filtering out nutrients from water, have to behave in some way to stay alive. Typically, this means that they have to forage. Similar to contragrade processes inside cells, neural processes like inhibition may also be energy costly, and may require motivation to be engaged in.

2.2 Oscillation, excitation, and inhibition

In this section, I will briefly explain excitation (increasing frequency) and inhibition (decreasing frequency) in the context of neurons. I will then go into how networks of oscillators can mediate specialization by means of how they are connected (topology), and how inhibitory connections can act as dynamic membranes that separate sub–networks from each other, analogous to how cellular membranes shield chemical processes. I will then introduce the important concept of *complexification* and multipolar causality. I will explain how interacting oscillators can support hierarchical processes via energetic coupling and resonance. Such interacting oscillators can have *emergent* properties. An example of complexification is then presented in the form of perceptual grouping, or gestalt processes in vision.

Oscillations, or cycles, are ubiquitous in the physical world. An everyday example of oscillations may be found in the electricity grid, where turbines like windmills or water wheels rotate magnets to generate oscillations of electricity. Oscillations are found everywhere also in biology. Examples include genetic oscillators that cyclically produce proteins inside cells, and chemical oscillators that turn the rate of chemical reactions up and down. In the nervous system, the primary source of oscillations is the neuronal cell, with its characteristic action potential spiking behaviour. Spiking is different from the even rotation of the water turbine in the electrical grid, and systems that generate this kind of accumulate–and–discharge behaviour are called *relaxation oscillators* since they tense gradually and then quickly relax. A leaky tap in the kitchen is an example of a relaxation oscillator, where water builds up inside the tap, until a drop falls. Inside the neuronal cell the situation is more complicated; electrical ions take the place of accumulating water, though what is released is typically neurotransmitters like *glutamate* or *GABA*.

Glutamate and GABA are in fact the most common *excitatory* and *inhibitory* neurotransmitters. Excitation of a neuron means that it will increase its rate of emitting action potentials. Using the kitchen tap analogy, increasing the water flow in the tap will increase the rate of drops that the tap produces, making the tap more excited. Turning down the water flow will in the same way reduce the rate of drop, yet this is not quite like inhibition. To understand how inhibition works in neurons, it necessary to know about positively and negatively charged ions, and also to know about *ion channels* that open and close based on voltage. For now, it is sufficient to know that to form action potentials, the internal voltage of a neuron must be high enough to open certain voltage gated ion channels that let in positively charged ions like calcium. If a neuron receives inhibitory GABA neurotransmitters, negatively charged chloride ions are channeled into the cell, which reduces the voltage. The calcium channels are thus less likely to open. We can modify the kitchen tap analogy to be closer to the neuron by installing shunting valves on the tap. Opening such a valve is like receiving a GABA signal, and will reduce the rate of drops that comes out of the tap, in effect inhibiting the emission of drops.

As mentioned earlier, neurons are always found in groups, and so excitation and inhibition only makes proper sense in a *network* of neurons. In an organism, neurons are connected to *sensors* of various kinds, that can tell the organism something about its environment, or about the state of the organisms own body. For example, if an organism like a small earth–worm has chemical sensors that increase their spiking activity dependent on the concentration of food, and other sensors that increase activity dependent on the concentration of toxic chemicals, those two signals can be connected to neurons that control the muscles of the worm. Together a small nervous system like this may be enough to orient the worm towards food, and away from toxic chemicals. It also makes sense for the food neurons to inhibit the toxic sensing neurons and vice versa. In this way, the worm will be less likely to get stuck trying to both approach and avoid the same spot, since there might be both food and toxic chemicals present at the same place.

To sum up, neurons connected into a network like this tend to naturally adjust their activity so that overall the whole network does not spend more energy than necessary. We will look closer at this process of energy optimization next.

2.2.1 Oscillation processes optimize energy dissipation

Oscillation is a *dissipative* process, and in this section, I explain the implications of energy dissipation on oscillation processes, and how they by thermodynamic necessity are bound to move towards dissipation minimums conditioned on energy availability.

All vibrational systems tend to reduce their energy dissipation to the minimum possible; this is called the *minimum energy principle* (Pauling, 1940; Petoukhov, 2016). This principle applies across systems from the smallest quantum mechanical, to the largest involving gravity waves (French, 2018). It includes also the oscillators making up slime moulds, and networks of neurons. According to Pauling (1940), resonance is an energy minimizing property, such that oscillators that can resonate tend to shift their frequencies accordingly. The same applies for *synchrony* (that is, phase–locking), and can be seen in the classic example of metronomes gradually coming to click both at the same frequency, and in synchrony.

The principle of minimum energy is what enables slime moulds to find optimal paths between food sources. The oscillators will tend to relax towards their least energy costly state, and stabilize there. However, environmental particularities and local topography will impose constraints on the minimization process. This means that even small changes can lead to different minimal paths.

As we have seen in the slime mould, networked oscillators come to be responsible for different functions in the larger system. Some of them are involved in chemotaxis and movement. These vary their frequency dependent on whether they are in contact with nutrients or repellents. Other oscillators shuttle nutrients around the network, making sure that both building materials and energy is available where needed (Vallverd'u et al., 2018).

2.2.2 Oscillation processes support complexification and hierarchy formation by means of coupling and resonance

Deacon (2011) mentions several examples of complexification in everyday experience, including vortices forming in flowing rivers, and Bernard convection cells forming in heated fluids and in the atmosphere. However, the music of an improvising jazz band is also an example of complexification, since the performance depends on the interaction between all the players. Likewise the behaviour of an ecosystem is an example

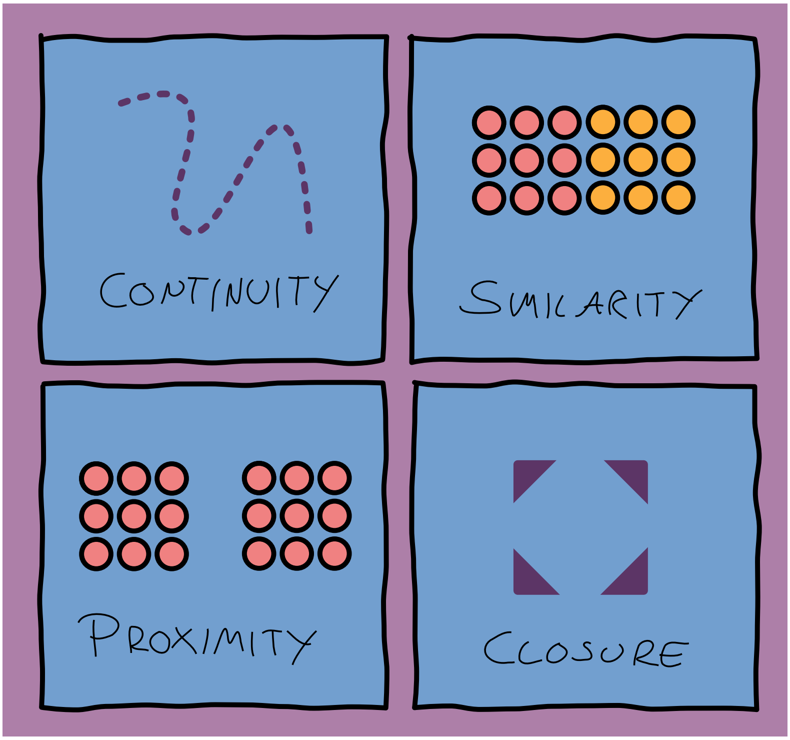


Figure 2.2: Gestalts are grouped percepts due to phase locked oscillations in the perceptual system. **Continuity** grouping makes discontinuous stimuli appear to hang together. **Similarity** grouping couples together similar stimuli, in this case circles. **Proximity** grouping makes stimuli that are spatially clustered appear to become a unit. **Closure** grouping induces the illusion of closed shaped in stimuli where an apparent shape is in front of another.

of complexification because all organisms, including plants and insects, will tend to interact with each other in a periodic fashion. This is not to say that the strength of the interaction is the same or is equally important for all organisms; organisms may also be subject to substitution to some degree.

According to Esghaei, Treue, and Vidyasagar (2022) dynamic coupling of different frequency bands provides a mechanism to functionally label and route signals from different sensory modalities.

According to Arnulfo et al. (2020) 100–400 Hz high–frequency oscillations (HFOs) may be synchronized between widely distributed brain regions. HFO synchronization is both transiently enhanced and suppressed in separate frequency bands during a response–inhibition task. The authors propose that the bursts of synchronized HF oscillations reflect the broadcasting and transmission of brief packets of information in large–scale neocortical networks.

Gestalt phenomena may be viewed as a class of complexification phenomena, also mediated by coupling oscillators (Keil, Müller, Ray, Gruber, & Elbert, 1999). Figure 2.2 shows a selection of visual grouping phenomena: continuation, closure, and proximity grouping. These phenomena may come about due to the topology of low level sensory neurons (Cara, Bettini, & Ursino, 2003). Models also suggest that, as the neurons lock phase and resonate according to their stimuli, sub–populations will spontaneously form due to the minimal energy principle (Wang, 1995; Wang & Terman, 1997). These populations may thus mediate the percepts of the various grouping types by their rhythmic behaviour.

C. M. Gray and Singer (1989) measured the activity of neurons in the visual cortex of cats, while showing them light bars aligned with their receptive fields. They observed that the neurons oscillated with a frequency around 40Hz. This frequency is in the *gamma band* of neural frequencies, which is typically defined to be from 20–70Hz, and is relatively fast for cortical neurons. The measured neurons were phase–locked, or rhythmically locked, to a *local field potential*. This field potential was recorded by the same electrodes used to register neuronal activity. A local field potential is the electric field which surrounds neurons, as distinct from the electrical activity of the neurons themselves. Interestingly, the researchers also recorded the activity of neurons in the dorsal lateral geniculate nucleus of the thalamus of the animals. These are areas that are closer to the retina in the signal path than is the visual cortex. The neurons recorded here did not show the same gamma band activity. According to the authors this appears to indicate that synchronization happens in particular in the cortex, but less so in sub–cortical areas.

Continuing work on cats, Engel, Kreiter, König, and Singer (1991) recorded neuronal activity from single neurons in an area of cat visual cortex that mediates perception of motion. The researchers showed moving bars of light to the cats in three different conditions. First they showed bars that moved in different directions, one up and one down. Then they showed bars that moved in the same direction but were discontinuous. Finally, they merged the two bars, showing one continuous moving bar. Results showed that gamma coherence was only present when the bars were moving in the same direction. Populations were most strongly synchronized in the final condition, when a single moving bar was shown. These results support the notion that visual percepts can be grouped by spatial continuity and when perceived to have the same direction and speed of movement, and that oscillatory synchronization is one mechanism by which this grouping can be mediated.

In sum, neurons connected in networks affect each other through their rhythmic activity, and automatically adjust their rhythms and phase so as to expend as little energy as possible. Neural populations can have complex rhythms with both slow and fast components, just as a drummer can combine slow rhythms on the kick and snare drums with fast patterns on the hi–hat. In the brain, the slow and the fast rhythms can mediate different types of function, such that slower rhythms can serve to keep different populations synchronized, while the faster rhythms can carry signals from sensors, and can also allow sensory signals to be grouped together.

2.3 Perception involves composition of statistical building blocks

Perception is necessary to be able to interpret the state of the environment and hence make sensory signals actionable. Depending on the needs of the organism, perception is used to interpret space, atmospheric pressure dynamics as sound, or vibration dynamics as tactility and touch. Some organisms are likely born with adequately functioning receptive fields, but in many complex animals like mammals, receptive fields develop over time. This development is typically based on the statistics of the gradient fields that make up the animal's sensory world. Examples of such fields are photon fields that are used for vision, gas pressure fields used for sound, or even magnetic fields used by birds for navigation (Ritz et al., 2009).

Hubel and Wiesel (1962) recorded activity from single cells in the visual cortex of anaesthetized cats. The cats were treated such that their eyes were open and stable, and their pupils dilated enough to adequately receive light. The researchers then exposed the cats to various simple visual patterns, like black and white rectangles rotated at different angles. They found that cells in the visual cortex, close to the pathway from the eyes, responded in a combination of excitation and inhibition, e.g. like exciting along a dark line, and inhibiting along the light borders. Other cell groups had the reverse pattern, exciting along the light borders and inhibiting along the dark line. The researchers referred to these as ”simple cells”, as in having responses to simple visual features like contrast differences and lines. Results of experiments indicated also that some cells responded to moving bars, rotated at specific angles. These were in turn termed ”complex cells”. These findings were largely replicated by Hubel and Wiesel (1968), using anaesthetized macaque and spider monkeys and recording again from the striate part of visual cortex.

Kobatake and Tanaka (1994) measured neuronal responses of cells in V2, V4, and anterior and posterior part of the inferotemporal cortex (anterior and posterior IT) of anesthetized macaque monkeys, while showing them three dimensional objects which were progressively simplified. These stimuli were compared with the response of the recorded neurons to simple black and white bars. Areas V2, V4, anterior IT, and posterior IT, are progressively further away from the retina. The results showed that cells recorded in V4 and anterior as well as posterior IT are sensitive to complex shapes. The authors also found that cells in V4 and posterior IT reacted to combinations of shape and texture, as well as shape and colour. Similarly, Gallant, Connor, Rakshit, Lewis, and van Essen (1996) showed polar and hyperbolic sinusoidal gratings to anesthetized macaque monkeys, and recorded the responses of 103 neurons in the animals' visual area V4. Again, this area is further away from the retina than the area recorded from in the experiments mentioned above, that recorded from neurons in striate parts of visual cortex. The results of these experiments showed that V4 neurons are sensitive to such visual features as curvature and simple shapes. Also the results indicated that neurons in this area tend to become active independently of how the stimulus is rotated.

Rolls, Aggelopoulos, and Zheng (2003) continued the work by Kobatake and Tanaka (1994); they recorded specifically from IT neurons in awake macaques that were trained to search for objects either in complex natural scenes, or with plain backgrounds. Results indicated that the size of receptive fields of the neurons changed dependent on whether objects were presented with a plain or complex background. In particular, the receptive fields appeared to be smaller when the complex background was used. According to the authors, these results indicate that IT areas can form unambiguous representation of targets, even if they are presented with complex backgrounds, like in natural scenes.

These papers together indicate that there is some gradient of complexity to which neurons respond, and that neuronal populations closer to the retina prefer simpler stimuli than do neuronal populations that are further away. There is also the question of how these preferences develop, which is related to how *acuity* develops. Acuity is in this context sensitivity to visual detail. Huang et al. (1999) accelerated the development of inhibitive neurons in mice by using transgenic method. In particular, they changed mouse DNA to over–express brain derived neurotrophic factor (BDNF) which is involved in the growth of dendrites and axons of inhibitive interneurons. During critical periods of learning, like in the development of the visual system, there appears to be a domination of excitation over inhibition (Blue & Parnavelas, 1983). In fact, these periods may come about because of the slower maturation of inhibitory circuits compared to excitatory ones. When inhibition balances excitation, the learning rate of neurons is reduced (Kirkwood, Lee, & Bear, 1995). In the experiments reported by Huang et al. (1999), the BDNF transgenic mice showed earlier maturation of visual acuity compared to control mice without the genetic modification. The authors thus propose that BDNF is involved in the development of the visual system. But it appears also to show that inhibition could be necessary for proper development of the ability to perceive visual detail.

In relation to this, Tjøstheim and Balkenius (2019) propose a way that visual perception might develop to first only represent coarse differences, but then gradually come to be sensitive also to fine detail . In this paper, we present a computational model of visual perception that supports both coarse and detailed receptive fields. The model network achieves this by hierarchical, or cumulative inhibition of neurons. In the model, neural units all tune to statistics of inputs. But the units are connected to progressively increase inhibition in a hierarchy. The outcome of this is that the least inhibited units get tuned to coarse patterns, while those that receive more inhibition progressively get tuned to more detailed patterns. As the receptive fields stabilize, they make up a collection of pattern building blocks that can be combined to represent common impressions in the environment.

This composition of building blocks is however not sufficient to regulate behaviour. It must be combined with the grouping dynamics discussed above to separate foreground and background sensations, and with attentional processes discussed in a later chapter. However, it must also work in conjunction with a memory system to support recognition, which can in turn contribute to biasing behaviour. We turn to this next.

2.4 Recognition requires both perception and memory

The use of statistical building blocks in perception means that an organism can form reusable patterns that afford recognition. Even if a stimulus is not exactly equal to what has been experienced before, the statistical nature of the receptive fields will ensure that there is a good chance that important stimuli will match, and recognition can happen.

In Balkenius, Tjøstheim, Johansson, and Gärdenfors (2018) we introduce a model of a memory system that includes a network for identification and recognition, as well as for spatial localization, and a working memory. In this section, I will focus on the identification network, and will come back to the localization- and working memory networks in a later chapter.

In the model, perceptual information (that is, the sensory building blocks described above) is fed into a fully connected network that functions as a content addressable memory. This means that given an input, the network will dynamically settle into a state that represents the most similar stimuli that is has learned before. In the artificial neural network literature, this is referred to as a Hopfield network (Hopfield, 1982), and can store stimuli by means of association, or Hebbian learning (Hebb, 1949).

In the mammalian brain, this recognition pathway is often referred to as the *what* pathway (Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983). In the context of vision, it includes the occipital cortex and the temporal cortex. The former of these corresponds broadly to the perceptual functions discussed above, while the latter corresponds to the content–addressable associative network that was used in the model.

Although not modeled in Balkenius et al. (2018), the actual learning of associations in biological systems is dependent on networks that represent value. Both negative and positive value will increase the learning rate of the associative network, since both can increase excitation in different parts of the brain. In practice this means that things that affect physiological processes in the organism will be learned by the network, and thus can later be recognized. The organism can then behave appropriately, including engaging in making choices.

2.5 Network motifs with flexible neurons

The slime mould is made up of oscillators, and shows how interacting oscillators can mediate allostasis, growth, and reproduction. However, multicellular organisms tend to use specialized cells to handle the brunt of cognitive processes. Neurons can act as oscillators in the sense that they can display stable frequencies, but that are subject to excitation and inhibition. However, neurons are typically more flexible in their behaviour than are the biochemical oscillators that make up the slime mould. In this section, I go into some of the details that afford neurons this flexibility, and explain how different network motifs can mediate cognitive fundamentals like recognition and choice.

Neurons can act as oscillators since they produce rhythmic action potentials. On their own, neurons may be of limited use; but like most cells neurons exist in tissues, and it is networks of neurons that in particular have flexible information processing properties. Neurons may be connected in a variety of ways, including by chemical and electrical synapses. In addition, intercellular space can act as a reservoir for modulator substances (see Braun and Leblanc-Sirois (2016) for a review). Some chemical synapses also support being dynamically gated on and off which makes for very fast and flexible routing of signals (Yang, Murray, & Wang, 2016). Neurons do not necessarily emit action potentials immediately when stimulated; internal electrochemical mechanisms are arranged such that stimulation must be sufficient to exceed a voltage threshold. After an action potential, a refractory period ensures that neuronal behaviour is rhythmic, since it must then pause as between beats on a drum. This rhythmicity allows neurons to have a very wide behavioural repertoire, from steady firing, to complex rhythms, via bursting behaviour (Buzsáki, 2006). Thresholds may also be employed to raise activity in some pathways, attenuating activity in others, but also to conserve resources by reducing activity. Finally, neuromodulators can affect the behaviour of neurons by affecting both bias voltages (Beaulieu & Gainetdinov, 2011; Neve, Seamans, & Trantham-Davidson, 2004) and learning rates (Lendvai & Vizi, 2008).

Neurons can link up in a variety of topological patterns that can support higher level cognitive processes. I will look at three common motifs here: mutual inhibition; association; and recurrency. First, the mutual inhibition motif entails that all participating neurons inhibit each others. This will result in all but the few most stimulated neurons being active. This motif is involved in selecting percepts (Baca, Marin-Burgin, Wagenaar, & Kristan, 2008; Blakemore, Carpenter, & Georgeson, 1970; Xiong & Chen, 2002) as well as in the process of choice (Gold & Shadlen, 2007). Second, the association motif involves bi–directional connection between neurons, such that the connection is strengthened between those neurons that are active at the same time. This is the Hebbian learning paradigm mentioned above (Hebb, 1949). Association networks bind together sensory streams into unitary groups, but are also used to bind value information to sensory streams (Carmichael & Price, 1995, 1994). Third, recurrent motifs entail also a form of Hebbian association, but in this case, self connections with e.g. varying myeliation can contribute to a neuron being stimulated by its own past output at several different times. This motif is likely used to mediate many kinds of sequence memory, including episodic memory (Neophytou, Arribas, Levy, Park, & Oviedo, 2021; K. L. Smith, Lee, & Swann, 1998; Whittington & Jefferys, 1994).

In paper 6 we proposed a model of location choice that included several of these motifs (Balkenius, Tjøstheim, Johansson, Wallin, & Gärdenfors, 2020). Choice between two spatial location was mediated by channeling excitation based on the perceived value of each location into a network that employed first mutual inhibition, then recurrent activation into leaky accumulator units. Hence choice is affected both of the perceived value of each location, but also of how long that location was attended.

I will attend to choice processes in some more detail below, but they will recur also in later chapters. The sense in which stimuli and choices may be considered meaningful is however of importance, and I will consider this next.

2.6 ”Meaning” as biological perturbation and behaviour

The word ”meaning” has several connotations in human terms. It may refer to purpose and positive affect, as in a ”meaningful life”, but it may also refer to semantics in terms of what the ”meaning of words” are. In this section, I will explain how what is meaningful to an organism may be usefully interpreted in the sense of semantics: as physiological affect on the organism, and as ”appropriate behaviour to engage in”. Finally, the appropriateness of behaviour may be understood as that which contributes to allostasis.

According to Cohen (2000), and Barandiaran and Moreno (2006) what is meaningful to an organism may be understood as the functional effect of the organism's interaction with its environment. The most primary way sensory impressions may acquire this form of meaning, is in direct physiological effect in terms of negative or positive affect. The pain of a piercing thorn to a paw, or of direct sunlight to an eye, are examples of negative affect. The pleasure of warming a cold body in the sun, or of sweet taste, are examples of positive affect. Stimuli that somehow induce stress in the organism, tend also to induce negative affect and engagement of the sympathetic nervous system. In effect, the sympathetic nervous system frees up energy from stores, inhibits repair processes, and makes the organism ready for action. Stimuli that reduce stress, including reducing physiological needs like hunger or thirst, tend to induce positive affect. Positively valenced signals attenuate the sympathetic nervous system, and excite the parasympathetic nervous system. The parasympathetic nervous system facilitates rest, growth, and repair processes, and enables energy stores to be replenished (McCorry, 2007).

The other way that stimuli may become meaningful to an organism is by means of the behaviour they motivate. For a predator, the smell or sight of prey can motivate stealthy approach; the sight of a rival can motivate aggressive approach. For a prey organism, the smell of a predator will motivate avoidance, or freezing behaviour. In both cases these stimuli will engage the sympathetic nervous system, channeling energy to the motor system.

In Tjøstheim, Johansson, and Balkenius (2019), we present a model that simulates aspects of familiarity based trust formation, using stimuli that could be considered meaningful in the way explained above . In this model, abstract stimuli representing soft and painful touch were paired with stimuli representing faces. The model contained also networks that simulated the effect of valenced stimuli on pupil dilation, thus connecting a social process like trust formation to involuntary regulation of the pupil. In a qualitative way, this model reproduced observations from experiments with human subjects where pupil dilation predicted trust in interactions.

2.7 Choice, discrimination, and generalization

The ability to make meaningful choices as discussed above, depends on recognition of what is going on in the external world. Recognition in turn is dependent on discrimination and generalization. That is, neuronal populations must be able to distinguish between stimuli that require different types of behaviour; if it is food, it should be approached and eaten, but if it is a predator, it should be avoided. Similarly, populations should generalize over stimuli that require the same behaviour; both an insect and a berry may be food, and both a hawk and an owl may be predators.

Hanson (1959) trained pigeons to respond to stimuli with different colours, such that one colour gave a food reward at variable intervals, while the other did not. The pigeons were divided into groups such that the non–reward colour had varying amounts of longer wavelength. After training they showed a range of colours to the pigeons and counted the number of pecking responses to each of these colours. The resulting distribution is called a *generalization gradient*. It shows how response is generalized beyond the original reinforcing colour, and is often shaped like a bell–curve with a central peak. The authors found that in their experiments, the peak of the generalization gradient was not centered on the reinforced colour, but was *shifted* to the left with a varying amount, dependent on far away the non–reinforced training colour was from the reinforced one. The closer the two colours were, the more to the right the peak was shifted.

Terrace (1963) used a similar setup to Hanson (1959), using red and green colours as training stimuli for pigeons, with red being rewarded at variable intervals, while green was never rewarded. Differently from Hanson, Terrace examined the notion of how ”errors”, in this case the selection of the unrewarded green stimulus, can be affected during training. In his experiment, he started with only the red stimulus; then he introduced the green stimulus, but with a different brightness and duration to the red one. The brightness and duration were then gradually made equal to that of the red stimulus. With these conditions, the results indicated that the pigeons were able to learn without making errors. Letting the red and green stimuli have the same brightness and duration from the beginning, resulted in the birds making errors. The birds that were trained in the presence of the green stimulus from the beginning also were reported to display fear behaviour to that stimulus, as well as an increased response to the positive red stimulus.

With human subjects, Lissek et al. (2010) studied how a diagnosis of panic disorder might affect perceptual generalization. The authors showed differently sized rings on a computer monitor to subjects. Across the spectrum of sizes, one extreme was associated with danger, the other extreme with safety. The researchers measured the startle–blink reflex using electromyography as an indication of a fear response. The startle–blink reflex typically consists in an eye–blink response to a surprising, or startling stimulus; electromyography entails measuring the electrical potential of muscles using conductive patches that can be stuck to the skin. Results from these experiments indicated that subjects with panic disordered generalized their blink response to rings with up to three units of size difference from the original danger stimulus. In control subjects without panic disorder, this generalization was limited to a single unit of size difference. The authors conclude that heightened fear responsiveness like in panic disorder can induce over–generalization effects to similar stimuli.

These studies on birds and humans show that biological systems can adapt to both generalize and discriminate the precision of stimuli that elicits behaviour. But what mechanisms exist at the neuronal level to mediate this broadening or narrowing of stimulus to behavioural connections? To elucidate this issue, Dudek and Bear (1992) conducted electrophysiological experiments on hippocampal brain tissue from rats. Using electrical frequencies between 0.5 to 50Hz, they stimulated neural populations of the slices, and measured the post–synaptic potential afterwards. They found that using pulses of 1–3Hz would consistently result in a reduction, or depression, of the post–synaptic potential. Conversely, if the stimulation was at 50Hz, and increase of the potential was measured, while at 10Hz no change could be observed. The researchers also noted that increase of potentiation could be prevented by using NMDA receptor antagonists. Antagonists bind to receptors without activating them, reducing the number of receptors that activating substances can bind to. NMDA receptors are excitatory, and depend both on voltage level, as well as glutamate stimulation, to open and admit positive calcium ions (Shadlen & Kandel, 2021b). According to the researchers, the experiments suggest that synaptic depression can be produced by long–term low frequency stimulation of NMDA receptors, while synaptic potentiation can be achieved by high–frequency stimulation that admits sufficient calcium ions.

These studies may be interpreted to shed light at both the psychology, and the neuroscience of behavioural triggering, in the following way. First, the behavioural experiments show that animals readily generalize cues that predict food, based on similarity. Second, this generalization is limited, or shaped by absences of reward. This mechanism can contribute to sharpen the recognition of cues, allowing animals to become more discriminate. Thirdly, the experiment on humans indicate that responses to danger can be particularly indiscriminate and general. This makes sense in an evolutionary context where erring on the side of caution tends to keep animals alive longer. Finally, the study on rat neurons provide at least one plausible mechanism for mediating generalization and discrimination, via long term potentiation, and long term depression, dependent on the frequency of NMDA receptor stimulation.

2.8 Recognition and choice as cognitive fundamentals

In this section, I will argue that recognition and choice may be considered fundamental cognitive processes. It should be clear from the above sections that these processes are not elemental in themselves, they are mediated in turn by neural behaviour; but they may be considered fundamental cognitive processes in the sense that they are the minimally necessary processes for an organism that is not continuously fed energy from its environment.

In this thesis I am using the notion due to Cohen (2000) that the presence of options, or alternatives can drive the need for cognition as something more involved than stimulus–response behaviour. Having options makes it useful to have processes that can *evaluate* those options, and to make *choices* based on the evaluations. Ecological niches vary in their requirements to make choices, and the options available to an organism depends much on that organism's behavioural repertoire. This again depends on its specific anatomical structure.

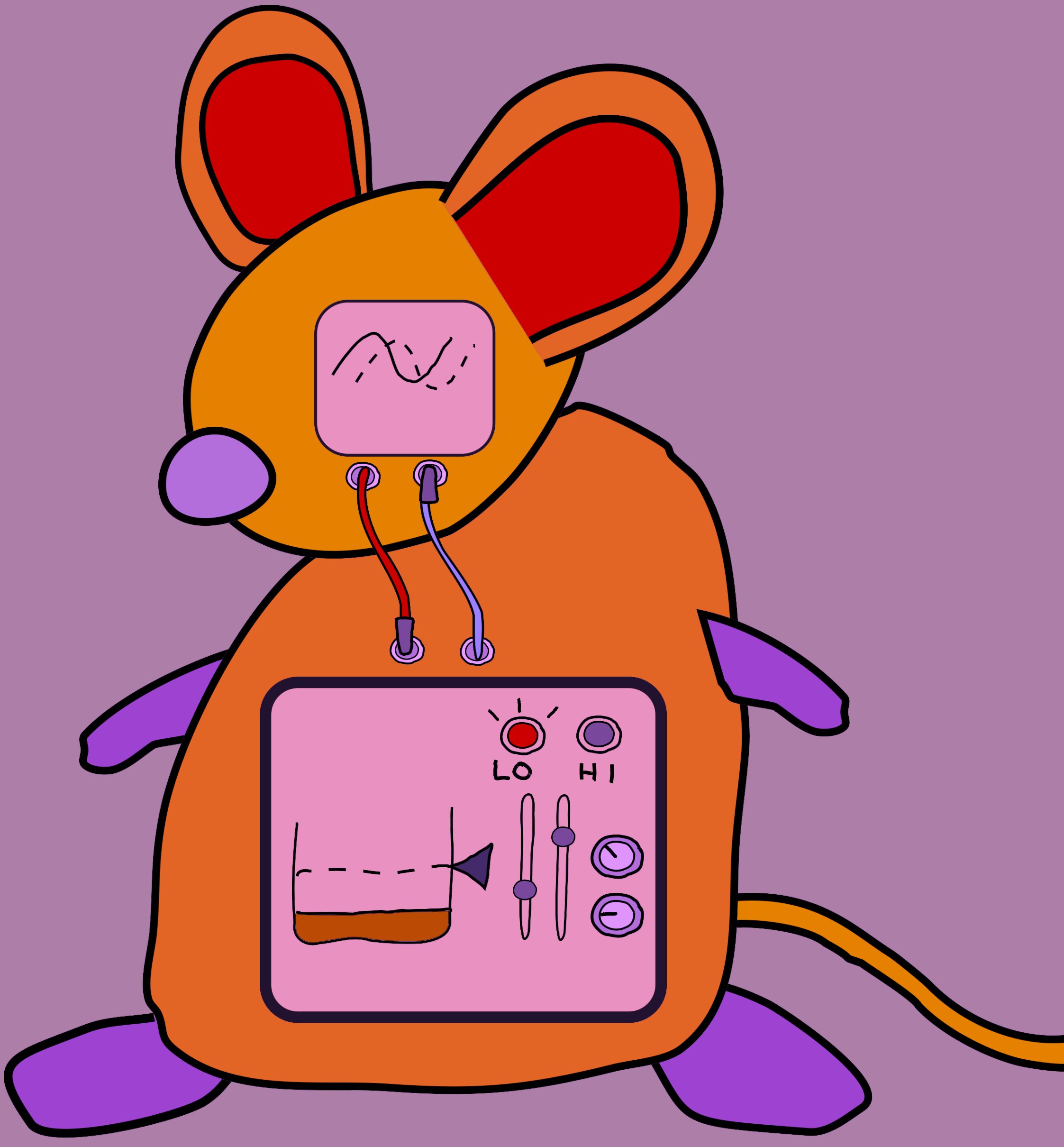
Let us consider again the slime mould. Does the slime mould have options, or can it make do with responding automatically to stimuli? To what extent can the slime mould go against the thermodynamical gradients that powers its oscillators? Takagi, Nishiura, Nakagaki, Ueda, and Ueda (2007) subjected slime mould to an obstacle in the form of a chemical repellent in a narrow corridor. They observed that after some "indecisive" waiting time, the slime mould either carried on through the obstacle, split in two and both progressed through and retracted, or completely retracted. A mathematical model based on reaction–diffusion dynamics apparently could reproduce this behaviour. Hence in a very simplified way, the slime mould has three options: to proceed over the obstacle or to retract, or to split and do both. From the model, the authors suggest that the indecision is likely due to internal instability dynamics of the slime mould tip. Thus depending on the exact dynamical configuration of the moment, one of the three behaviours will win out. To some extent this may be similar to contragrade movement, limited by the comparative strength of the repellent, versus the growth dynamics of the slime mould.

Now, let us briefly consider a selection of cognitive processes, and how they contribute to recognition and choice. First, perception represents that which might be recognized, and recognition may be understood as a dynamic filter of perceptual information. Second, valuation may be considered as contributing to choice; value information is typically that which biases choices of goals and behaviours. Third, memory processes may contribute to both recognition and choice. I will consider four types of memory and attempt to indicate their contributions. Perception memory consists of receptive fields and thus contribute to recognition; sensory association memory also contributes to recognition in that associations enable processes to be recognized via any of several sensory streams; episodic memory can contribute both to recognition e.g. when recognizing situations, but it may also carry valence information from remembered outcomes and thus affect choice; working memory may contribute to both recognition and choice since it can be used to do short term comparison, but working memory may also be used to hold rules that can affect the choice of goals and behaviour. Finally, executive function in the sense of impulse inhibition and prediction can contribute to recognition in the form of biasing perception, and also regulate choice in the sense of maintaining approach to non–immediate goals.

For the purposes of this section, I distinguish behaviour as well as goals as that which is chosen. Goal directed mental behaviour like working memory manipulation and abstraction are cognitive processes, but here I lump such processes together with motor behaviour. Hence in the context of this thesis I identify recognition and choice as *fundamental* cognitive processes in that I consider them to be what cognition is about: the control of behaviour; perception, valuation, memory, and may be considered *elemental* cognitive processes that are necessary to realize recognition and choice.

2.9 Summary

In this chapter, I have traced an arc from thermodynamics to cognitive processes. We have seen that thermodynamical gradients can spontaneously induce shape–patterns in liquids via Bérnard cells and vortices. In biological neuronal cells, differences in concentrations of ions inside and outside the cell is the basis for action potentials. Action potentials are a type of oscillation, and neurons are a type of oscillator. Linking oscillators together, like in the central nervous system, will automatically cause them to spend as little energy as possible, by means of resonance and synchronization. Resonance and synchronization in turn give rise to grouping processes in sensory networks, such that patterns that are close together, or of same size, or colour, may be perceived as belonging together. Visual perception works by composing small patterns like light and dark into larger percepts, that can be recognized by memory networks. Networked biological neurons display extra flexibility compared to basic thermodynamical oscillators, since they have thresholds, and are susceptible to modulation by a range of transmitter substances. These substances can both increase and reduce the rate of action potentials of a neuron. For an organism, perception and behaviour is deeply intertwined, such that the ”meaning” of percepts can be interpreted as what kind of behaviour is associated with it. In this sense, the meaning of a recognized predator is to run away, and the meaning of internal hunger signals is to go look for food. As organisms grow in complexity they typically have larger behavioural repertoires, and with this comes the necessity to make choices. The ability to recognize the meaning of sensory signals, and to choose suitable behaviour from a range of options may thus be considered fundamentals of cognition, and also one of the contours to which this work pertains.


LO
HI

Chapter 3

Allostasis and regulation

Allostatic processes are dependent on regulation, which implies that cognition too is dependent on regulation, while also being a means for regulation. Feedback control is thus a fundamental part of biological systems: it makes sure that the organisms stays within allostatic boundaries. In this context, the cybernetics framework is a useful framework for understanding cognition.

In this chapter, I start with informational free energy, and what it means that organisms must minimize informational free energy to maintain allostasis. After that I delve more into how feedback contributes to regulation, and the specific roles that positive and negative feedback plays. Then I explicitly connect allostasis to regulation, and explain how the former implies the latter. A foraging animal is subject to perturbations from its environment, and feedback processes contribute to coping with such perturbations. But some perturbations can be beyond the operating range of an organism, pushing it away from homeostasis and into pathology. Finally, noise and thermodynamic entropy will also exact their toll on living systems, degrading their regulatory processes over time.

3.1 Informational free energy

Thermodynamic free energy can be formulated mathematically as the difference between two probability distribution, and this formulation also holds for information. In this section, I will explain the relationship between thermodynamic– and informational free energy, and how the concept of information may be usefully understood in the context of biology and allostasis.

For the purposes of this thesis, the relationship between thermodynamic free energy and informational free energy, is that the former is required to reduce the latter. More specifically, according to Friston et al. (2006) the free energy at any one time is the discrepancy between what the internal model of the brain says should be happening, and what the sensory apparatus says is actually happening. Discrepancies between the two are called *prediction errors*. Further, the brain's expectation signal is used to inhibit the sensory signal, such that only the difference between the prediction and the sensory signal is propagated (Wagner & Smith, 2008). Hence when something doesn't turn out as expected, this difference signal is relayed through the perceptual system, until eventually it can affect which choices are made. The point is that prediction errors can be reduced, or minimized, at any of these stages. For newborn animals, whose perceptual system is adjusting, the prediction errors induce the formation of receptive fields. In some sense, this is what the model we presented in Tjøstheim and Balkenius (2019) does. Although this model does not explicitly use prediction and prediction errors, the structure of the model allows it to be modified to do so, and hence is an avenue for further work. However, once the perceptual system is mature, predictions tend to be about situational outcomes, and could be e.g. about the behaviour of other animals. Note that the predictions I talk about most in this thesis are not so much about the external environment, but instead concern the state of the body, as described in the previous chapter. In these cases, what is required is not so much the update of the internal model of the world, but behaviour that results in the ingestion of food.

When the internal model does get updated in the cortical areas of the brain, the difference between the backward flowing model signals and the forward flowing sensory signals is computed by superficial pyramidal neurons in layers 1-3 (Bastos et al., 2012). The feedback signals from the internal model are mediated by neurons in layer 5 and 6. There are then three timescales involved: that of glutaminergic ion channels, of modulatory ion channels, and most slowly of adaptation of synaptic weights. Hence when an animal moves in the world, frontal areas continuously compute the expected pattern of action potentials at the different levels of the the sensory hierarchies (Bastos et al., 2012). Whether prediction errors are minimized by means of synaptic updating or foraging behaviour, thermodynamic free energy in terms of available ATP will have to be expended to do so.

In the next section, we will encounter the concept of *cybernetics* and feedback control. This is a mathematical framework for understanding how processes can be regulated. After that, we will see how allostatic processes can be viewed from this perspective, and that allostasis implies regulation.

3.2 Feedback processes, regulation, and discretization

The term cybernetics alludes to the Greek word for a person who steers ships, and was first used in Wiener (1948). The essence of cybernetics is *feedback*, and how using feedback can either stabilize systems, or force them out of a stable state and into a new one.

In the context of navigation, it is easy to see how a helmsman keeping a straight course towards a port must continuously adjust the rudder of a ship to compensate for wind and currents. The helmsman may use his eyes to look at a landmark and make adjustments depending on whether the bow of the ship is pointed at the landmark or not; or he may use a compass and make sure the compass is pointing in the right direction. Using the terms of cybernetics and control theory, the eyes of the helmsman are the *sensors* of the regulated system, while the helmsman's arms and the rudder of the ship makes up the *effectors*. Furthermore, the landmark is usually called a reference value, or a *setpoint*. The difference between the bow of the ship and the landmark, is called the *error*. Hence the regulated system, consisting of both the ship and helmsman, uses the error to continuously adjust the rudder via the effectors. This minimizes the error.

The same principle applies in biological systems. The body temperature of a human should typically be close to 37 degrees Celsius. This is an example of a setpoint; if the body temperature drops too low, quick muscle contractions, or shivers, work to return the body temperature to the setpoint. Hence the muscles are effectors, working to reduce the error measured by temperature sensing neurons in the body.

In the nervous system, predictions can be understood as types of setpoints too, while prediction errors corresponds to regulation errors; effectors include both the activity of neural populations, the modulation of activity by neuromodulators, as well as synaptic adaptation involved in learning (Ha & Cheong, 2017; Haenisch & Bönisch, 2011; Lin & Kuo, 2013).

The examples above have been about *negative* feedback, that is feedback that is subtracted from the setpoint to yield an error. However, *positive* feedback is also used in the body. Typically, positive feedback is used in cyclical processes, to make sure the end of a cycle is reached. The most important example of positive feedback for this thesis, is that of the action potential elicited by neurons. When an action potential is formed, the buildup of charge in the neuron at one point reaches a point of no return, where additional ion channels open up and the voltage ramps up to a maximum (Hodgkin & Huxley, 1952). The process after the point of no return is thus an example of positive feedback, and how it can mediate discrete either–or conditions. In the next section, we will look at how prediction error information is a crucial element of allostasis.

3.3 Allostasis implies regulation

One of the fundamental reasons why biological systems need to be regulated is due to a natural growing disorder in all such systems. We encountered this concept, called *entropy*, in chapter two. Without proper regulation, organisms become increasingly less orderly. For example, cellular compartments can break down, and with them, biochemical processes like protein formation, or the synthesis of ATP. Once this process gets going the whole organism will in the end break down and die.

Here, we will look at biological processes from the cybernetic perspective, and identify some regulated systems within neurons, and in neural populations. One of the most important ones for this thesis is the regulation of calcium ion flow. Above we saw that calcium is important in synaptic adaptation for long–term–potentiation, and long–term– depression. As it turns out, calcium is also important for the health of neurons as such. Yoon, Won, Ryu, and Gwag (2003) reports that neurons are subject to tight regulation of calcium flow, since insufficient flow of calcium ions will trigger the mitochondrial death process and terminate the neuron. However, calcium flow is again dependent on stimulation from presynaptic neurons, via the opening of NMDA receptor channels. This means that a single neuron cannot upregulate calcium flow on its own. One possibility for understimulated neurons, is to upregulate their input from populations that represent sensory detail, such that stimulation can come from prediction errors of detail information (Ahissar & Hochstein, 1997).

Similarly, calcium flow must be limited on the upper end, since excess calcium will also kill the neuron (Connolly & Prehn, 2015). The upper and lower limits on activity that a neuron can handle, as well as the need of neurons to be stimulated to stay alive, impose interesting boundary conditions on neural tissue that are related to *information*. As we shall see in the chapter on foraging, these constraints are a way for bottom up mechanisms to influence the behaviour of the organism as a whole.

At the behavioural level, feedback has a slightly different intuitive meaning, as when children get feedback from their teachers in a class. Feedback on behaviour can also come in educational games, and Tärning et al. (2020) present a theoretical model how feedback can fail in this setting. According to this model, students need to both notice and process feedback information before they can make sense of it and finally adapt their behaviour. In the framework presented in this thesis, these steps may be mapped to attention, in that students must notice that the feedback information appears; perception, in that students must assemble visual signals and read text; recognition, in that students must recognize which behavioural options the feedback entails; and finally it must affect choice of behaviour. This can include cognitive behaviour, such as recalling earlier event sequences and reinterpreting them. Compared to low level physiological regulation processes where attention and motivation do not play a part, high level interactive processes are at once much more complex but also more interesting. The principle of feedback regulation is nevertheless the same; the intricate process of interacting with the game will contribute to shape the students knowledge towards an ideal target.

The plasticity involved in biological systems, albeit requiring strong regulation, can also make them robust and forgiving of environmental changes, as we will see next.

3.4 External and internal disturbances can push allostasis towards its limits

In the previous section, we saw how biological systems can function in diverse environments, and under a range of environmental conditions. But organisms routinely get sick and die. What happens regulation fails, and biological processes are compromised? Environmental perturbations that challenge allostasis are usually called *stressors*, and stress and its effects are thus the subjects of this section.

Let us go back to our helmsman and his ship, and let us assume that he is steering a sailing ship where everything is manual. What happens if the ship enters a strong current which the helmsman didn't know about? The sailing ship will veer off course, and the helmsman may try to compensate by sharply turning the ship in the opposite direction. An overshoot then happens, meaning the ship now is heading too much the other way. Again he tries to compensate, but the sharp turn of the rudder combined with the current again leaves him off course. The ship is now cycling, or oscillating back and forth between extremes, but never quite reaching the steady course it had. The upshoot is that the helmsman is having to spend much more energy than before. What happens if he gets tired, and can no longer sufficiently compensate? The ship will then eventually point in the same direction as the current, and go wherever it leads.

The same principle applies to physiological processes in the body, and even to neural processes in the brain. Many people find that their metabolism changes as they get older, and they gain weight. Some then try to reduce their body weight by sharply cutting down on food. They may find that they do indeed loose weight, but that an increased appetite compels them to eat more. The result is the familiar yo–yo effect, where no lasting weight reduction is achieved. In this case, the metabolic changes due to aging is like the ocean current, and the steering behaviour of the helmsman is like dieting. In the end people are unable to muster the energy for dieting, and the body will sail its course.

The brain goes through natural cycling as the body tracks e.g. the day and night cycle by means of the circadian system. But in some cases, often after stressful life events, systems in the brain can be pushed out of their regular rhythms, and into oscillation. Manic depression may be an example of this. The exact mechanism behind this condition is not well understood, but some researchers argue that irregularities in calcium allostasis may be involved (M. J. Berridge, 2013). Others theorize that energy mechanisms and problems with neuronal mitochondria are likely causes (Mansur, Lee, McIntyre, & Brietzke, 2020). In the latter case, the authors suggest that individual variations in circadian system, and hence the ability to handle disruptions to the day–night cycle, could contribute to pushing an individual's metabolic system out of sync with the body's energy needs. These changes could then manifest as symptoms of depression and insufficient energy availability, alternating with mania and an excess of energy.

There are many changes in an organism's environment that can cause stress. For animals, lack of food is common, and dry seasons with little water is stressful too. Humans can experience severe stress when they are forced to change their habits after loosing a job. Hence, both forest animals and humans may be compelled to change their foraging habits, and their allostatic behaviour, due to stressors.

Some of the negative effects of stressors depends on the degree to which they are surprising. The necessity for prediction in allostasis is the topic we turn to next.

3.5 Allostatic regulation requires prediction

Organisms typically alternate between being asleep and awake. They alternate between eating and defecating, between foraging for food, and mating. Bodies have internal rhythms, and are sources of patterns. Above I use the example of a helmsman on a ship that starts to oscillate when entering a strong and irregular current. The oscillation occurs because the helmsman is forced to *react* to the current. That is, some time would elapse between the helmsman noticing that the ship was moving away from the setpoint, and him managing to turn the helm. However, if there was a repetitive pattern in the current, he could learn to *predict* it, and turn the rudder early enough to compensate for the change he anticipated. With enough experience the helmsman would be able to keep the ship heading in the correct direction without oscillation.

Physiological processes in the body is analogous to the regularly shifting current, while the central nervous system typically has the job of the helmsman. Concretely, an animal becomes hungry sufficiently long before it actually needs energy so that the blood sugar stays within allostatic range. Another example is a student becoming nervous before a presentation, but becomes calm once the presentation actually starts. Some of the nervousness consists of the stress response of the body regulating up the availability of energy in anticipation of meeting the demands that the presentation will make. If the demands are accurately predicted, the student will become calm as the presentation proceeds, since there is enough energy available, and that energy is made use of according to predictions.

To be able to make adjustments in the body like this, an interoceptive network involving the subcortical *hypothalamus* and areas of the cortex called *insula* and *cingulate cortex* together act as the helmsman, using both targeted neuronal signals to areas in the brain, but also broadcast messages to the body in the form of hormones (Glinoer, 1997; Pagotto, Marsicano, Cota, Lutz, & Pasquali, 2006).

According to Craig (2002), the hypothalamus is part of the *interoceptive pathway* that projects information from peripheral parts of the body to the brain. One endpoint of these projections is the *insular cortex*, placed in between the temporal lobe and the parietal lobe. Lovero, Simmons, Aron, and Paulus (2009) conducted MRI scans on human subjects while applying a touch stimulus to their hands after the subjects were told by voice that this would happen. The imaging data showed that mid–to–posterior insular regions were activated when the touch stimulus occurred, while the anterior insula was activated as the subjects anticipated the stimulus after being cued.

The perception of threat and danger is accompanied by signals from the body, like increased heart rate and faster breathing. Critchley, Wiens, Rotshtein, Öhman, and Dolan (2004) instructed human subjects to judge the timing of their own heartbeats, while scanning them using fMRI. The authors report measuring increased activity in the subjects' insula, as well as their somatomotor– and cingulate cortices. The accuracy of the subjects in the interoceptive task correlated with measured activity in right anterior insula and the surrounding opercular cortex. There was also a correlation between the grey matter volume in these areas, and subjects' reported visceral awareness, and accuracy in the task. Finally, accuracy was also correlated with indices of negatively valenced emotional experience.

Holtz, Pané-Farré, Wendt, Lotze, and Hamm (2012) investigated which areas of the brain were particularly activated in relation to the anticipation of threat symptoms like hyperventilation. Before putting them in an fMRI scanner, the authors used a task in which human subjects were guided through hyperventilation behaviour, and where cues were used to enable the subjects to anticipate hyperventilation or normal ventilation. They recorded the subjects' reported level of fear of unexplained bodily sensations. At this point, the researchers found that subjects that reported high level of fear, also reported higher level of bodily symptoms, and showed higher potentiation of the startle reflex. The subjects were then put in the fMRI scanner, and the same cues were presented to the subjects as in the hyperventilation task. However, the task itself was now not conducted. The results from this experiment showed that the anterior insula was activated in the subjects as they received the cue for hyperventilation, along with the orbitofrontal cortex, the dorsal anterior cingulate cortex, and the dorsomedial prefrontal cortex. Subjects that reported higher levels of fear had an overall stronger activation of the previously mentioned areas, and this activation tended to persist even when they learned that actual hyperventilation would not occur. See Barrett and Simmons (2015) for a review and a theoretical framework concerning prediction in allostasis.

Tjøstheim et al. (2019) employed a biologically inspired computational model of pupil dilation connected to soft or painful touch signals via a simplified interoceptive pathway, to investigate how interoceptive signals can influence familiarity based trust, as well as approach and avoidance behaviour. The model was exposed to signals representing human faces in two sequences, first while receiving signals representing gentle or painful touch, then without any such stimulus. The model reproduces pupil behaviour associated with positive and negative signals, and predicts that negative interoceptive signals have an asymmetric negative affect on trust. It also predicts that negatively valenced encounters dominates positive ones, and hence tend to overall motivate avoidance when positive and negative signals are equal.

In summary, these studies indicate that regulation of the body's internal environment involves prediction, and that areas particularly involved in this aspect of allostasis include hypothalamus, insular cortex, and cingulate cortex. There are also indications that interoceptive signals like pronounced heart beat or ventilation is negatively valenced and associated with threat, and can motivate avoidance behaviour.

In a regulated system, correct engagement of effector is not only dependent on information from sensors being accurate, but also that control information itself is not distorted on its way to effectors. In the next section we will look how both of these processes can degrade over time and how this may affect a cognitive system.

3.6 Noise and entropy degrade regulatory processes over time

Recall from the first section of this chapter that thermodynamic processes are intrinsically noisy, and that biological processes are therefore noisy too. In this section, I will explain how noise and entropy adversely affect allostasis over time, and that this is one major contribution to the aging process.

Once again let us consider our helmsman aboard his sailing ship. Perhaps the ship made it out of the current, but new challenges await. The ship now enters an area of magnetic rock formations, and a thick fog descends. We need also to take a closer look at the ship's rudder mechanism. The rudder on this ship is namely connected to the helm, that is the ship's steering wheel, by means of iron shafts and gears. It so happens that iron is elastic and can be twisted; repeated high force on the iron rods can also fatigue them making them less able to transfer the movement of the helm to the rudder. We remember also that the helmsman did a lot of turning earlier, coping with the ocean current.

Two sources of uncertainty has now entered the regulatory system of the ship; firstly, the helmsman can no longer see the landmark he was following quite as well, and the magnetic compass has become less reliable due to those magnetic rocks. Secondly, he can no longer be completely certain how well his turning of the helm actually transfers to the rudder.

In the brain, the process of aging can impose similar uncertainty both on senses and effectors, as the fog and the unreliable rudder in the ship example. Accidents, injuries, and social stress can make demands on the body's energy system, making it less able to maintain itself. One particularly important factor appears to be the very fundamental aspect of mitochondrial reproduction when cells divide. According to Lane (2005), cells start out with pristine mitochondria, able to produce ATP at high rates. Mitochondria have their own DNA, apart from that of the host cell. In the same way that copying errors can affect cellular DNA and produce mutations, such error can affect mitochondria and produce mutations there. Unfortunately, most mutations affect mitochondria negatively, and reduce their efficacy. That means that, as time goes on, the capacity of cells to produce ATP decreases, and hence decreases also those cells' means to power all the biochemical processes that are required to clean up waste, and maintain cellular structure (Stefanatos & Sanz, 2018). We can observe the sum total of this process of reduced metabolic capacity and increasing entropy as aging.

One consort of aging is Alzheimer's disease, which in particular appears to affect the medial temporal areas involved with memory and spatial cognition. In the next section we will look into how pathological processes such as this, though dreadful, may still provide a window on cognitive processes.

3.7 Pathology as a window on cognition

Dementia is an outcome of several pathological processes, including Lewy body disease, Parkinson's disease, and Alzheimer's disease (Raz, Knoefel, & Bhaskar, 2016). Here I will focus on Alzheimer's disease because one of the characteristics of this process is impairment of medial temporal areas that are involved with spatial navigation, and the use of cognitive maps (Coughlan, Laczó, Hort, Minihane, & Hornberger, 2018).

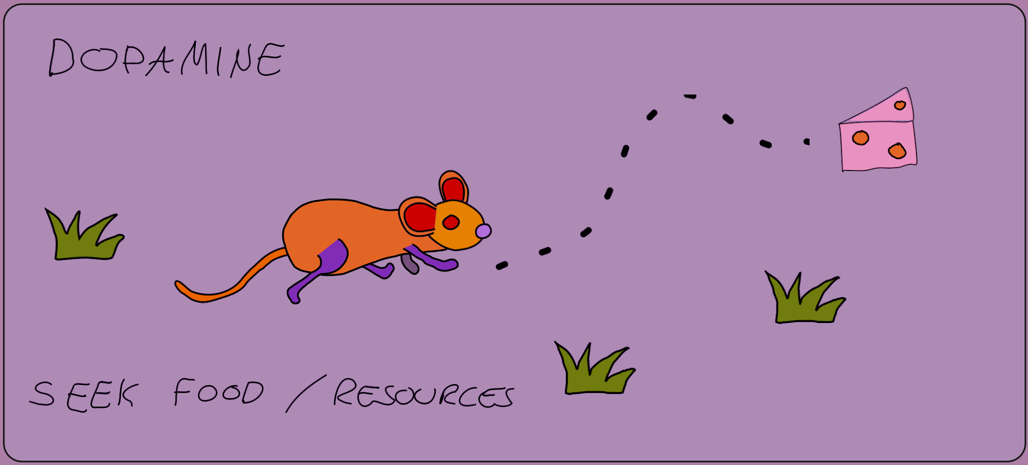
As reviewed by Serino, Cipresso, Morganti, and Riva (2014), people in early stages of Alzheimer's disease tend to experience more problems with navigating complex environments. In particular, they note that Alzheimer's disease appears to make the use of mental maps more difficult. These situations occurs when you cannot directly move towards the place you want to go to, but must construct in your mind a way around a building or something that stands in your way. The term for this way of finding your way is using an *allocentric reference frame*, where *allo–* means "other", in the sense of "other than yourself". By contrast, when you can just walk right up to your goal, you are using an *egocentric reference frame*, or the viewpoint of your own eyes. Serino et al. (2014) also note that as neurons die in the spatial cognitive areas, shifting between the two reference frames becomes more troublesome. Hence in Alzheimer's disease, the recognition of where you are in relation to places around you, can become more difficult. As a consequence, your choices of how to navigate to your goals become more limited.

For people experiencing psychosis, recognition of what is happening in the world also becomes challenged. Kapur (2003) theorizes that some element of psychosis is due the dopaminergic system marking things as salient when they shouldn't be. However, dangerous things that should be avoided, like predators, must also be salient. In both cases, wrongful labelling of percepts in this way will motivate approach or avoidance behaviour that can be maladaptive. For example, stalking behaviour may result from misattribution of people as salient for approach, e.g. as a romantic partner (Miller, 2012), while paranoid avoidance and hostility may come from marking people as salient for avoidance or aggression (Tiernan, Tracey, & Shannon, 2014).

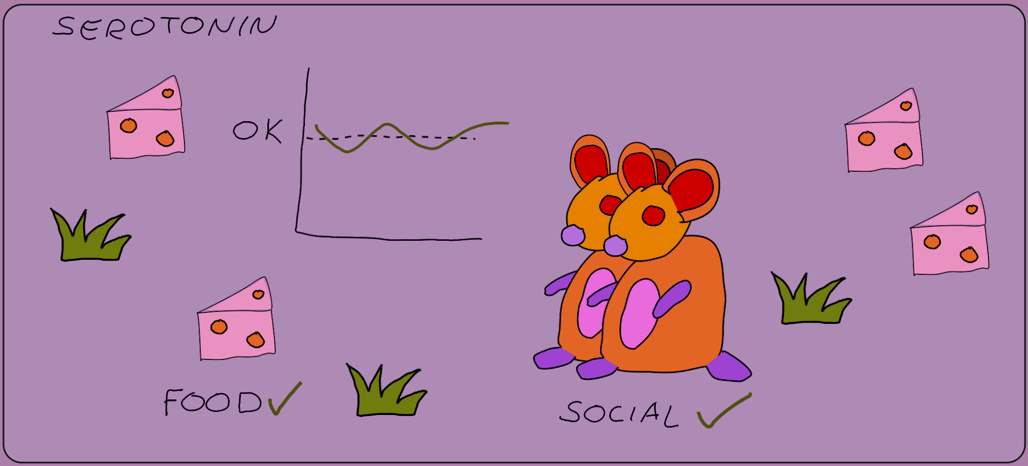
The mental models that are formed from these kinds of misattributions of salience, are often referred to as *delusions* (Gerrans, 2014). Understanding how they are formed, not only has the potential to increase understanding of cognitive processes in general, but brings also hope that the suffering experienced by the people involved can be reduced. Gerrans (2014) presents the hypothesis that delusions may be the product of the *default mode network* constructing subjectively adequate causal chains in response to unexpectedly hypersalient stimuli. The default mode network is typically conceptualized as being active ”by default” when people are not engaged in goal directed behaviour (Buckner, Andrews-Hanna, & Schacter, 2008). Although it connects together several areas, the important function in relation to delusions, is that it can make up possible trains of thought that can be interpreted as possible causal paths (Gerrans, 2014). In the absence of dysfunction, momentary default trains of thought, such as ”he looked at me, he’s out to get me” are quickly dismissed by prefrontal areas that can dampen default activity based on more general beliefs about society and the state of the world (Gerrans, 2014). But if these prefrontal areas are somehow weakened or damaged, default thoughts can come to influence choices and hence behaviour.

3.8 Summary

In this chapter, I have introduced the concept of allostasis, which is the fundamental biological process of maintaining life. Allostasis implies that an organism is regulated by means of feedback processes. Feedback ensures that physiological variables like blood–sugar levels, or the concentration of salt in the body, are neither too high or too low. However, rectifying allostatic errors requires energy, in the same way that keeping a sailing boat on course requires energy and muscle power. Both external and internal perturbations can challenge the allostatic system: going out into cold weather without clothes increases the energy the body has to spend to keep itself at 37 degrees, while a virus infection increases energy demands by the immune systems. To keep physiological variables within their boundaries, it is not enough for the body to blindly react to the environment; rather it has to be one step ahead, and predict what is going to happen. Thus, the muscular system will start working in preparation before you lift a heavy bag of groceries. You will feel hungry some time before the energy from the food is needed, so that it is already digested and ready at the appropriate time. But in any physical system there is noise, and the same is the case in the body. Noise from internal heat will affect both sensor signals from eyes and ears, and from sensors that measure blood sugar. Likewise there is noise in the effector signals that are tasked with rectifying regulation errors. During the aging process, the noise in the system tends to grow, which means that the system as a whole becomes more and more likely to fail. Studying failure in the body and the nervous system, can provide valuable information about how these systems work. For example, during the course of Alzheimer’s disease, people tend to experience growing difficulties with using mental maps, and have trouble understanding their own spatial position in relation to their surroundings. This knowledge may give insights into both the causal underpinnings of Alzheimer’s disease, but also into spatial cognition.


DOPAMINE
SEEK FOOD / RESOURCES


OPIOIDS
EAT & LIKE FOOD


SEROTONIN
OK
FOOD ✓
SOCIAL ✓

Chapter 4

Energy, allostasis, homeostasis

Neural processes, and therefore cognition, are constrained by energy availability and mitochondrial capacity at the subcellular level. Like cognition, behaviour may be considered an allostatic process, both in terms of foraging, but also in terms of directly regulating energy conservation and dissipation through e.g. physical activity.

Allostasis is the process of maintaining life, and is concerned with making sure that sufficient energy is available, and that available energy is balanced with energy cost. The main point of this chapter is that organisms primarily engage in allostatic processes, and that cognitive processes can therefore also be considered from the perspective of allostasis. That is, we can ask what the is role of cognition in allostasis.

Several neuromodulators systems are involved in allostasis, and I will present three of them here, discussing them in the context of energy management. I argue that the serotonin system mediates long term sampling of environmental suitability; the dopamine system mediates approach to resources; and the opioid system mediates energy conservation which, along with energy gain, is the basis for positive value.

Allostasis happen at multiple scales, and in the final section I begin at the neuronal level, with calcium homeostasis, and work up to the organismal level with the tension between novelty and boredom.

4.1 Allostasis as energy management

Allostasis is the process of working towards homeostasis. In biological systems, the organism is seldom in actual homeostasis and if so, only for short periods of time. This is because perturbations from the environment tend to be rife, resources within the organism tend to deplete quickly, and waste products build up. However, allostasis is primarily about *energy management*, as I will endeavour to explain in this section.

If you stand in the middle of a forest, see bushes moving, and hear the rustle of leaves of something approaching, you may feel uncertain, both about what is in the bushes, but also about what to do. It could be a bird or a cat, but it might also be a bear. This uncertainty will induce stress, and depending on the how big and dangerous you estimate the thing to be, your body might prepare to run away and divert energy to muscles, and away from the brain. If the thing sounds small and harmless, energy might instead be diverted to networks in the brain that can reduce uncertainty by matching to experience or to semantic knowledge. This might be a benign example, but a more serious situation could e.g. involve the loss of a job without being able to easily replace it. Uncertainty will then regard how to be able to pay bills, feed a family, or even have somewhere to live.

If uncertainty is not reduced, the brain may go into a persistent energy crisis which can contribute to malfunction not only in the brain, but also in cardiovascular– and digestive systems (Peters, McEwen, & Friston, 2017). In cases like this, the *allostatic load* increases. This term can be understood as referring to how much energy is required by the body to maintain allostatic processes, and keep physiological systems within the healthy zone. If the allostatic load is persistently too high, it can induce pathological processes including impaired memory, diabetes, and clogging of arteries Peters et al. (2017).

There are several ways the cognitive system can deal with uncertainty to prevent the increase of allostatic load. First, control of attention can reduce uncertainty not only by attending directly to areas in the sensory field that are unclear, but also by seeking out information actively. This option will invariably entail learning, which has the effect of increasing predictive power. In the example of finding a new job, these two alternatives might involve searching on job sites on the internet, and calling up companies to learn who might be hiring. However, a third way to reduce allostatic load is *habituation*, or adjusting expectations. Hence if no jobs are to be found, expectations may be shifted from being employed with a high salary, to being unemployed and having less money. Goals may also shift from making a career, to making do, and focusing on family.

Hence energy management can and do happen at many levels at the same time; a stressful event can upregulate signaling substances that facilitate learning when the stress is over (Peters et al., 2017); similarly learning can also be facilitated when effort is made to cope (Pratt & Kelley, 2004). At the behaviour level, both approach and avoidance behaviour can be triggered, either to engage with a situation and reduce uncertainty actively, or escape from it if that is possible.

In the next section we will look closer at some of the neuromodulator systems that are involved in allostasis.

4.2 The neuromodulator systems that support allostasis in foraging

In the central nervous system, neurons are affected not only by excitation and inhibition, but also by several different substances that modulate both plasticity and spiking behaviour. Some of these neuromodulators are particularly important for allostatic behaviour like foraging, and in this section, I will present three of them. First, the *dopaminergic system* is involved primarily in aspects of motivation and energizing approach to goal–sites important for allostasis. Second, the *opioid system* mediates liking and hedonic aspects of reward. Since opioid receptors appear to be inhibitive and can be found on GABAergic interneurons, I will go into some detail arguing how the opioid system might mediate value by means of saving metabolic energy. Third, the *serotonergic system* performs a wide variety of functions in both the nervous system of the gut, as well as in the central nervous system. Overall though, the serotonergic system appears to keep track of long term environmental suitability. The other neuromodulators, that include acetylcholine, oxytocin, noradrenaline, and histamine, are important for learning, social cognition and social reward, and arousal. These neuromodulators are involved in allostasis as well, but with an exception of acetylcholine and learning, perhaps less involved in allostasis in the context of foraging.

4.2.1 The dopaminergic system: seeking, approach to allostatic goal–sites

The dopaminergic system complements the serotonergic system. The overall function of the dopamine system can be understood as mediating behaviour that brings a system closer to allostatic goals. This includes navigating to goal–sites that contain necessary resources like food or water, but also social goal–states, as well as informational ones having to do with curiosity and knowledge.

Dopamine is associated with a plethora of functions in the central nervous system, but is often associated with *reinforcement learning* (Glimcher, 2011; Watkins & Dayan, 1992), where those behaviours that tend to yield rewards are increasingly triggered at the presentation of a salient signal (Holroyd & Coles, 2002).

Researchers have been debating whether dopamine is directly associated with reward, referred to as *liking*, or whether it functions as a signal of unexpected reward, i.e. as *reward prediction error*, and that this signal is used to facilitate *learning* (Wise, 2004). There are also indications that dopamine is involved with perception via *incentive salience* that is associated with *wanting* (K. C. Berridge, 2006; K. C. Berridge & Robinson, 1998).

Dopamine appears also to be involved with *motivation*, the facilitation or energizing of behaviour. Matsumoto and Hikosaka (2009) found that when exposing monkeys to both rewarding as well as aversive stimuli, recordings of dopaminergic neurons in areas near basal ganglia, became active in response to both kinds of events. Hence dopamine is not only used in connections with reward, but also with negatively valued events.

Building on Alcaro, Huber, and Panksepp (2007) and others, Wright and Panksepp (2012) argue that the dopaminergic system can be understood as contributing to a brain system that mediates *seeking*, in the sense of seeking out and approaching resources.

Although evidence tend to support dopamine being involved with wanting, rather than liking, the opposite is the case for the opioid system which we turn to next.

4.2.2 The opioid system: positive value as energy savings and –gain

Where dopamine mediates triggering of allostatic seeking behaviour, the opioids appear to mediate the biological substrate for consummate value. In this section, I discuss how the opioid system might perform this function by regulating metabolic cost in the central nervous system.

The opioid system appears to be strongly linked with consummatory positive valence, in particular the value of food. There are three types of opioid receptors, called mu, kappa, and delta (Satoh & Minami, 1995). The mu opioid receptor is mainly inhibitive and expressed subcortically; the kappa opioid receptor is implicated in reward from fatty food, and in satiation, and appears to oppose mu receptor activity. The delta opioid receptor is implicated in rewards from fatty food, like the kappa receptor. Similarly to the mu opioid receptor, it appears to have a mainly inhibitive effect, but is expressed more broadly in cortex rather than subcortical areas (Satoh & Minami, 1995). According to these authors, opioid receptors are primarily disinhibitive, tending to reduce activity of inhibitive GABAergic interneurons.

Peciña and Berridge (2005) and K. S. Smith and Berridge (2005) investigated the effect of administering sugary rewards to rats, and observed the animals' stereotypical behavioural reactions. For example, rats tend to increase eating, and stick out their tongue in response to sweet tasting foods. For bitter tasting food, they shake their head and turn away. By using microinjections to the nucleus accumbens shell in the basal ganglia, of a substance that binds to mu receptors and hence tend to increase the behavioural response to sweet taste, the researchers found that the neurons that respond to sweetness are present only in a small area of the n. accumbens shell structure. Using the same method for the ventral pallidum area, the researchers found that ventral pallidum generation of increased food reward and increased eating behavior is related, but can be dissociated. The point is that opioid mediated liking, or hedonic impact, is closely associated with consummatory behaviour, in contrast to the seeking behaviour that we saw associated with the dopaminergic system. It is also interesting that both sugars and fats reliably induce liking, these being both very energy dense foods.

As mentioned above mu and delta receptors have an inhibitory effect on neurons, while kappa receptors oppose this inhibitory effect (Satoh & Minami, 1995). Typical physiological effects of opioids include suppressed breathing, drowsiness, as well as constipation (Merrer, Becker, Befort, & Kieffer, 2009). Hence mu and delta stimulation can reduce activity of smooth muscle, both in the respiratory system, and in the gut. Opioids also reduce arousal.

In a recent review, van Steenbergen, Eikemo, and Leknes (2019) propose that mu receptors can contribute not only to liking mechanisms in the basal ganglia, but also to networks involved in higher level decision making. The authors mention that such areas as orbitofrontal cortex, basal ganglia, the amygdalae, anterior cingulate cortex, and prefrontal cortex are now found to be dense in mu receptors. The authors propose that subjective value of reward may be increased also by these receptors reducing aversive arousal.

In summary, the opioid system may viewed as responding to energy gain through consumption of energy rich foods such as sugars and fat, but in some way also may mediate energy *saving* by lowering arousal, and tuning down energy costly inhibitive interneurons.

4.2.3 The serotonergic system: long term sampling of environmental suitability

The serotonergic system is involved in digestion as well as affective mood. Evidence suggests, however, that the overall function of the serotonin system is to sample the energetic suitability of an organism's environment. In this section, I discuss how this entails both sampling the availability of food, but also the availability of social support.

Serotonergic signaling is highly prevalent in the nervous system of the gut, the enteric nervous system (De Vadder et al., 2018). Serotonin is thus involved in metabolism, and participates in how food material is conducted through the digestive system. Since serotonergic activity contributes to the rhythmic contraction of the gut, a deficit of serotonin here can result in constipation, while excess serotonin signaling can induce diarrhea (Camilleri et al., 2002). In the central nervous system, serotonin is synthesized from *tryptophan* in the raphe nuclei in the brain stem. Outward projections from these areas modulate brain stem structures receiving meal–associated interoceptive and systemic signals (Tecott, 2007). The body cannot make tryptophan on its own, and is reliant on food sources to supply of it. The hypothalamus, which is one of the key areas involved in allostasis and the regulation of energy balance (Shadlen & Kandel, 2021a), receive serotonergic innervation and express serotonin receptors in all its component networks.

Reduction of serotonin synthesis, e.g. from a lack of tryptophan, is associated with increased motivation to eat, or *hyperphagia*. The opposite can reduce eating, inducing *hypophagia*, and also tends to increase metabolic rate (Donovan & Tecott, 2013). Serotonin is typically distributed by axons that diffuse it into extracellular space. It binds to receptors that are not part of synapses, but rather distributed along various parts of neurons (Braun & Leblanc-Sirois, 2016). A mechanism of *reuptake* continuously channels excreted serotonin back into axons to be reused. In experiments on mice, the absence of such serotonin reuptake increased extracellular serotonin concentrations, and made the mice less active than controls. The converse was the case when the researchers increased the reuptake rate by genetically increasing the expression of serotonin transporter proteins. That is, mice with higher reuptake were observed to run around more and were more active than control mice.

The serotonergic system appears to interact closely with the oxytocinergic system, which contributes to the mediation of social behaviour. In particular, interaction between these two systems seem to be necessary for social behaviour to be rewarding, and for such behaviour to be reinforced by reward (Dölen, Darvishzadeh, Huang, & Malenka, 2013). That is, oxytocin alone appears not to be sufficient to account for the rewarding effects of social interaction, but is dependent on serotonergic activity to trigger rewards. Conversely, serotonergic neurons in the raphe nuclei express oxytocin receptors, and oxytocin activity appears to facilitate release of serotonin (Yoshida et al., 2009).

Taken together, and from a foraging perspective, this evidence may be interpreted as the serotonergic system providing a long term average of nutritional quality of the environment. In concert with oxytocin, the serotonergic system reflects also the level of social support. Nutritional quality is related to which kind of food contains tryptophan, which includes among others seeds and nuts, as well as milk, fish, and poultry

(Nongonierma & FitzGerald, 2015) (but see Soh and Walter (2011) for evidence that nutrition alone is not sufficient to increase tryptophan levels in blood). These are all nutrient–rich foods, and in particular it appears that tryptophan is an indicator of protein sources (Nongonierma & FitzGerald, 2015).

4.3 Behaviour as allostatic process

Usually, allostasis is associated with the internal physiological processes that continuously regulate the organism, such as maintaining blood pressure, keeping blood sugar stable, and making sure the body is not overwhelmed by invading pathogens. In this section, I will explain how overt behaviour may also be considered primarily allostatic. I will focus mainly on foraging behaviours, but also make a point of discussing information– and novelty seeking behaviours that may not be as obviously allostatic.

Inside the body, regulatory loops have the privilege of being closely coupled. The peripheral nervous system with sensory cells can synapse directly with hypothalamic neurons. These can again synapse with epicrinal cells that can excrete hormones, or other effector substances, into the body. In contrast with such tightness, the coupling between an organism and its environment tends to be more tenuous. Bruineberg and Rietveld (2014) describe this looser coupling by means of an organism's "grip" on environmental affordances. This grip entails that the organism can interact with its environment, and draw resources from it, but also that it can adequately sense what is going on in it. Bruineberg, Kiverstein, and Rietveld (2018) makes the point in this regard that organisms need not necessarily be attuned to the "truth" of what is actually the case in its world, only that its sensory impressions are sufficient for it to maintain grip, and keep itself alive. As the authors point out, a perfect model of the environment would be of little use if it did not adequately indicate to the organism what behaviour it needs to engage in to maintain allostasis. An example of this could be that, smells and taste indicate to an animal, both where to go to look for food, and whether what it has found, is suitable to eat. If the smell is attractive and the taste is sweet, this could go a long way, both in telling the animal what it needs to know, and making it engage in eating behaviour that maintains supply of energy.

It can be useful to employ mathematical idealizations of information processes to increase understanding of animal foraging. Friston et al. (2015) use such mathematical perspectives to build a framework for how organisms can predict their environments. In the ideal case, an animal would have learnt all there is to know from an environment when there is no longer any difference from what it anticipates will happen, and that which actually happens. At this point, the animal is in a sense free to exploit the knowledge it has acquired, using it to maintain allostasis without having to pay metabolic costs to update the internal model with new information. In practice, this ideal case is not likely to come about very often. But it offers a useful perspective on so called *epistemic foraging*, also referred to as *active inference*. This can be interpreted as behaviour animals engage in specifically to learn more about their environment.

Since it is difficult to know much about animal's subjective states, it is a matter of speculation to confer on them human–like feelings of curiosity. But as reviewed by Gottlieb, Oudeyer, Lopes, and Baranes (2013), curiosity is associated with patterns of gaze behaviour, and these behaviours may have some overlap with those found in animals. The reduction of uncertainty appears to be one aspect of epistemic behaviour, and which also applies to gaze movements. Gottlieb et al. (2013) proposes that having an intrinsic motivation to reducing uncertainty by learning, would help organisms in the sense that their long–term evolutionary fitness could increase in rapidly changing environments.

Peters et al. (2017) couple uncertainty with metabolic cost via the concept of allostatic load. Recall that allostatic load can be understood as the energy cost involved in maintaining allostasis. In the context of curiosity, uncertainty about the environment may upregulate motivation for an animal to move around its environment, precisely in those places where it is unsure about what is going on. When it obtains information about those places, uncertainty will decrease, as will metabolic cost associated with this uncertainty. This energy saving, or relief from ongoing costs could then yield positive valence similar to the *liking* mechanisms described above.

.

4.4 Allostasis at multiple scales

Allostasis happens both at the scale of subcellular biochemical processes, and at the behavioural scale. In this section, I will expose the innards of neurons to explain the role of calcium in neuronal integrity, then move up into the level of networks to explain how plasticity can support cellular calcium homeostasis. Finally I will zoom out to the behavioural level and explore how boredom and novelty–seeking might be explained in terms of neuronal populations working to regulate themselves.

4.4.1 Neuronal allostasis and calcium homeostasis

Like the body as such, individual neurons must work to keep themselves in homeostasis. But neurons are sensitive to the activity of the neurons they synapse with. An active neuron both receives and exudes ions, and calcium ions in particular. Both

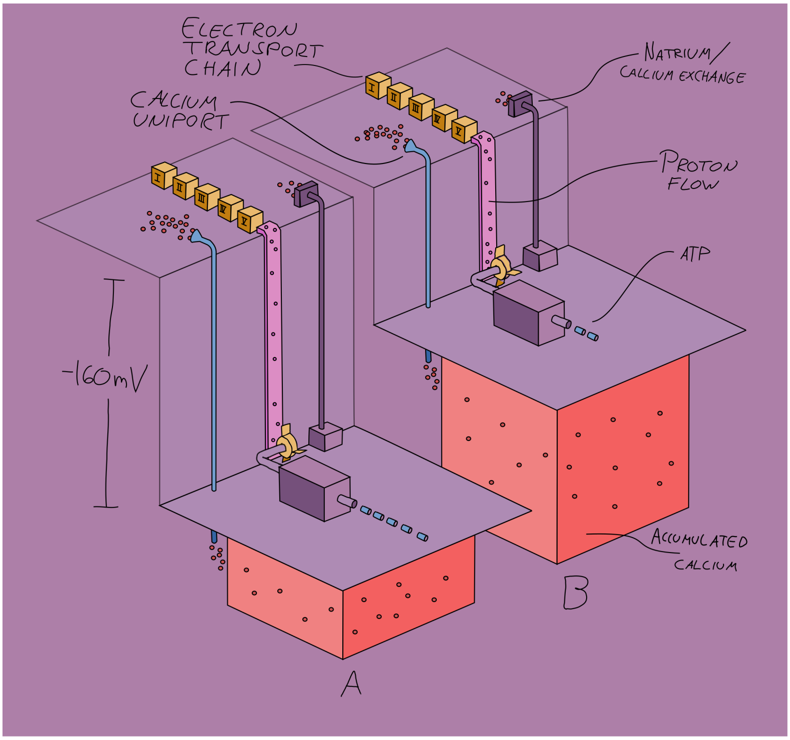


Figure 4.1: Calcium homeostasis in mitochondria. **A** When calcium flow is normal, the membrane potential inside mitochondria is around -160mV. **B** excessive calcium inflow reduces the mitochondrial membrane potential, which means that less power is available for making ATP.

too much, and too little calcium can stress neurons, and calcium dyshomeostasis is especially dangerous.

Too much stimulation by excitatory input can induce cell death in neurons, and is referred to as *excitotoxicity* (Choi, 1992). This process is typically associated with glutamate receptors, since glutamate is the primary excitatory neurotransmitter in the central nervous system. Choi (1985) showed that in cortical neurons, excess glutamate stimulation was followed first by swelling of cell bodies and dendrites, and followed by cell disintegration. According to the author, the first phase was associated with inflow of natrium ions, while the latter phase was dependent on availability of calcium ions. The notion that excess calcium ions can cause damage inside neurons was supported by (Hyrc, Handran, Rothman, & Goldberg, 1997), that showed via video–microscopy that stimulation of NMDA receptors in particular predicted cell death. In neurons, glutamate excitation is associated with two kinds of receptors, AMPA and NMDA. The AMPA receptor is typically activated first, since it is only dependent on glutamate stimulation, and admits natrium ions. NMDA receptors are dependent both on voltage level as well as glutamate stimulation, and admits calcium ions (Shadlen & Kandel, 2021b). Using cultured hippocampal cells, Stanika et al. (2009) showed that a pathway exists where calcium overload can affect mitochondria, and reduce their ability to produce ATP. Thus, excitotoxicity appears to be associated with excess level of glutamate stimulation, and over–activation of NMDA receptors, that then contribute to dangerously high levels of calcium ions inside the neuron; this again contributes to energetic stress growing where biochemical processes are impaired due to a lack of energy; ultimately this can lead to cell death (Connolly & Prehn, 2015).

As is usually the case in physiology, both upper and lower limits exist. This is also the case for neurons and calcium flow. Ikonomidou et al. (1999) injected seven day old rat pups with a substance that blocks NMDA receptors, and hence severely downregulates the amount of calcium ion inflow that neurons receive. In control rats, natural neuronal cell death was less than 2% of total neural density, while in NMDA blocked rats, the density was reduced by 12–26%. The authors noted also that neurodegeneration could be observed after as little as four hours. Ethanol binds to NMDA receptors as an antagonist (Mirshahi & Woodward, 1995), and so does nitric oxide (Sattler et al., 1999). This means that neuronal cell death in fetuses may also happen due to alcohol consumption during pregnancy (Chandrasekar, 2013), and even due to the use of anaesthetics like nitric oxide during birth (Ikonomidou et al., 2001). Similarly to Ikonomidou et al. (1999), Yoon et al. (2003) injected seven day old rat pups with the same NMDA antagonist and measured the amount of degenerate neurons, getting analogous results. To elucidate the processes involved, they removed cortical tissue from 14 day old fetal mice, and exposed that tissue to the NMDA antagonist. Analysis of the cells indicated that NMDA blockage affects mitochondria negatively. Calcium deficiency specifically activates a pathway in mitochondria involving the protein caspase–3, which coordinates destruction of several cellular structures. Once activated, this pathway ultimately results in mitochondrial death, which again will eventually kill the whole neuron by means of apotheosis.

In summary, these experiments suggest the existence of a physiological range of calcium concentration inside neurons, beyond which neurons are in danger of dying. That is, neurons can die both of calcium starvation, and of calcium excess.

4.4.2 Regulation of neuronal activity supports calcium homeostasis in mature brain networks

At the network level, inhibition and modulation are important ways that groups of neurons can affect how excited they become, and thus can regulate their concentration of calcium. The regulation of excitation is often referred to as *gain modulation* (Carandini & Heeger, 2012; Ferguson & Cardin, 2020). Perhaps the most important mechanism of gain modulation is inhibition via GABAergic interneurons. Borg-Graham, Monier, and Frégnac (1998) recorded neurons in the primary visual cortex of paralysed cats and kittens while showing them flashing images of optimally oriented bars. Despite simulations showing that such exposure should increase neural conductance by up to three times, the authors recorded only an increase of 5–20%. This much smaller response was found by them to be due to activation of inhibitive GABAergic synapses on the recorded cells. The authors refer to this as *shunting inhibition*, claiming that it works in a non–linear divisive way, rather than a linear subtractive way. Although later results (Ferguson & Cardin, 2020; Holt & Koch, 1997) indicate that the divisive effect appears to be due to a combination of both subtractive inhibition and multiplicative excitative stimulation, the outcome of division is to produce *normalization* of excitative inputs (Heeger, 1992).

This normalization effect has been shown in several model animals, including insects and mammals. Olsen, Bhandawat, and Wilson (2010) recorded responses from olfactory neurons in fruit flies. By using both stimuli unique to an olfactory neuron, as well as several general stimuli that all can yield a response, they showed that adding general stimuli to unique ones resulted in a normalizing effect on the olfactory neurons. Similarly to Borg-Graham et al. (1998), Carandini, Heeger, and Movshon (1997) performed extracellular recordings of simple cells in the primary visual cortex of anesthetized macaques. They showed images of grating patterns to the animals, with a variation of orientation and of visual detail. The results from the animal experiments were used to fit a mathematical model, which showed that cortical cells display non–linear effects when high contrast stimuli is presented to them. The non–linear effect could be accounted for by adding normalization to the model. Later, corresponding results have been shown for primary audio cortex (Rabinowitz, Willmore, Schnupp, & King, 2011) and parietal areas (Louie, Grattan, & Glimcher, 2011). See e.g. Carandini and Heeger (2012) for a review of normalization processes. These experiments show that combining inhibition with excitation can make sure that neurons are not excessively stimulated, and therefore can help maintain their dynamic range, and avoid saturation.

Above I discussed some neuromodulator systems that contribute to allostasis, including the serotonergic system and the dopaminergic system. Here I will present some ways that these substances can contribute to population level allostasis, and will also include noradrenaline. First, serotonin acts on several sensory networks, including for vision, audition, and somatosensation. Serotonin appears generally to modulate network activity downwards, that is to have an inhibitive effect. Ebert and Ostwald (1992) showed this in the cochlear nucleus of anaesthetized rats. They placed recording electrodes into this area, and exposed the rats to pulses of sound while injecting it with a serotonin solution . The results indicated that serotonin reduced spontaneous activity by about 70%, and the tone induced activity by around 30%. With bats under light anaesthetics, Hurley and Pollak (1999) recorded activity of serotonin affected neurons in the inferior colliculus in response to audio stimuli. They found that neuronal response depended on the type of signal; for tone bursts, the response was reduced, but for tone sweeps, some neurons were inhibited while others were excited, compared to controls. Costa, Kakalios, and Averbeck (2016) blocked serotonin re–uptake in behaving macaque monkeys. In a visual discrimination task, they observed that this blocking tended to slow down reaction times for the monkeys, and also made their perceptual performance worse.

Dopamine often works complementary to serotonin, so it might be no surprise that where serotonin reduces activity, dopamine can increase it. Zaldivar, Goense, Lowe, Logothetis, and Panzeri (2018) administrated L–DOPA to supragranular areas of primary visual cortex of anaesthetized rhesus monkeys, both in a diffuse way, as well as with focal injections. They found that the diffuse infusions increased activity, whereas the focal injections failed to do so. Phillis and Tebēcis (1967) and Albrecht, Quäschling, Zippel, and Davidowa (1996) both recorded activity of relay neurons in dorsal lateral geniculate nucleus of the thalamus of anaesthetized rats, by extracellular means. This entails inserting an electrode in tissue and recording electrical activity, and can be compared to intracellular techniques that measure activity in single cells. The recordings in these studies indicated that dopamine D1 receptors inhibited the relay neuron activity, while the D2 receptors excited it. Govindaiah and Cox (2005) found the opposite with intracellular recordings, but using brain slices from the dorsal lateral geniculate nucleus, of rats and mice rather than live specimens. These apparently contradictory results may possibly be reconciled if the injections on the live animals affected dopamine receptors on inhibitive neurons (Neve et al., 2004). What these experiments show, is that dopamine can have complex effects depending on which receptors are expressed on cells, and which cell types are investigated.

The effects of noradrenaline (NA) are similarly complex and depend highly on which receptors are expressed, as well as the concentration of noradrenaline. There are three main groups of noradrenaline receptors: A1, A2, and beta types. The A1 group typically have excitative effects at high levels of NA, while the A2 group are inhibitive and work at low concentrations (Ramos & Arnsten, 2007). Beta receptors appear to work similarly to A1 receptors, increasing excitation at high concentrations Ramos and Arnsten (2007).

In dorsal lateral geniculate nucleus of thalamus of anaesthetized rats, Rogawski and Aghajanian (1980) found that NA facilitated activity, and that this facilitation was blocked by A1 antagonists. They also found that beta antagonists did not have this effect. In the primary visual area of anaesthetized cats, Kasamatsu and Heggelund (1982) found that neurons affected by NA had improved signal to noise ratio, meaning that the neurons showed more contrast between their default activity when not stimulated by visual input, and when a visual stimulus was presented to them. A similar effect was observed by Foote, Freedman, and Oliver (1975) in awake squirrel monkeys. When administrating NA to the monkeys' primary auditory cortex via iontophoresis, extracellular recordings showed increased signal to noise ratio there. Iontophoresis is a technique where substances can be transferred through skin by applying local electrical fields.

One way to interpret these results, is that NA can have *contrast enhancing effects* in sensory networks, suppressing low intensity noise and amplifying signals (Ramos & Arnsten, 2007). See e.g. Jacob and Nienborg (2018) for a review of neuromodulator effects on sensory networks.

In summary, calcium homeostasis in neurons is mainly maintained by inhibitory regulation, that keeps neurons from becoming damaged by excess excitation. Neuromodulators are used diffusely to up– or downregulate activity in sensory networks in the case of serotonin, as well as to produce more local effects of respectively contrast enhancement for noradrenaline, and bias for dopamine.

4.4.3 Boredom and novelty

To understand the attractive and aversive qualities of novelty and boredom, we first have to go back to the notion of *prediction errors* introduced in the previous chapter. That is, we have to understand the role of predictions and prediction errors on network activity. Much theoretical work has been done on the role of prediction errors in the brain. According to theory proposed by Friston (2005) and Friston (2010), prediction signals are typically generated in limbic and paralimbic cortices, and propagate outwards to sensory cortices. This is based on a model proposed by Mesulam (1998). Predictions are mediated by inhibitive neurons such that when predictions are correct, signals from sensory neurons are damped, while errors are not. In this way, errors are causes of excitation in sensory networks. Since predictions are to a large extent based on experience (Hassabis, Kumaran, Vann, & Maguire, 2007), novelty will induce prediction errors, and can therefore increase excitation.

Uncertainty about sensory perception tend to increase activation too. Making use of the serial dependence bias, where the perception of features or objects is systematically biased toward recent sensory input, van Bergen and Jehee (2019) did fMRI recordings of human observers reporting on the orientation of stimuli presented in sequence. They decoded probability distributions from population–level activity, and found that sensory integration appear to follow a strategy where uncertain sensory integration is given less weight. The important point in this context is that the decoded probability distributions reflect the subject's uncertainty. Building on this, Geurts, Cooke, van Bergen, and Jehee (2021) tested a variety of computational models against empirical data from psychophysics– and fMRI experiments. In the latter experiment, human participants were presented with oriented gratings, and were asked to report on their observed orientation, as well as how confident they were of their judgement. The authors found that there was a correspondence between participants' reported confidence and the precision of decoded probability distributions. Further, probability distributions appeared to be mediated by populations in the insula, anterior cingulate, and prefrontal cortex. Both of these studies indicate that uncertainty in sensory data entails an increase of activity, involving increased number of neurons. It is therefore likely that uncertainty increases excitative activity in the brain, increasing also energy demands. This result is interesting, since we have seen earlier that insula and anterior cingulate areas are particularly involved in interoception and energy regulation.

As Raffaelli, Mills, and Christoff (2017) point out, boredom is somewhat hard to operationalize, even if it can be intuitively easy to recognize. In attempts to make aspects of boredom clearer and more precise, Goetz et al. (2014) conducted experiments on university students using an experiential sampling method. The participants were equipped with digital devices which prompted them at random times to fill in questionaires about affective state and their present context. The authors projected the results to a two dimensional space consisting of a valence dimension, and an arousal dimension. Using statistical methods to identify clusters in data, five areas were identified, being named in order of rising arousal: apathetic, indifferent, calibrating, searching, and reactant. The indifferent type was mildly positive, while the rest were identified as negative, with apathetic being the most negative followed by reactant; calibrating and searching were noted as equally and less negative than the two others. The high arousal, negatively valenced reactant boredom was associated with anger, and a strong desire to escape or avoid the boredom–inducing context. In contrast, the apathetic type of boredom appeared to be associated with a depression–like low arousal.

As predicted by calcium homeostasis in neural populations, too low levels of activity may be in need of rectification, as well as too high levels. In correspondence with the types identified by Goetz et al. (2014), calibrating, searching, and reactant boredom may all reflect growing intensity of desire to increase stimulation in humans. Bench and Lench (2019) conducted experiments on human participants viewing boredom inducing, repetitive images. The experiment included three conditions in which options were given to choose novel images. In the first condition, the boredom inducing images were neutral, and the novel images were displaying negative valence. They observed that subjects were more likely to elect to view the novel negative images when reporting higher levels of boredom. In the second condition, subjects could choose novel images that had both positive and negative valence, while also reporting their desire for novelty. The results showed that, subjects reporting higher levels of boredom, reported also greater desire for novelty. Greater desire for novelty also tended to correlate with a higher likelihood of choosing negative images. In the third condition, the researchers mixed both positive and negative boredom inducing stimuli. The novel images in turn displayed a range of affective valence, from negative to positive, but with higher intensity. In particular, novel choices were of same emotional valence as the boredom inducing images, as well as more negative and more positive. The results from this condition indicated that boredom could be induced by both positive, negative, and neutral stimuli. Further, subjects experiencing positive repetitive stimuli were more likely to choose negative novel stimuli.

Using questionaires administered to undergraduate students, Carriere, Cheyne, and Smilek (2008) found that there appeared to be a correlation between lapses of attention, cognitive errors related to attention, and memory failures on the one hand, and boredom and depression on the other. Depression is often associated with being over–stimulated, overwhelmed, and unable to cope (Bianchi, Schonfeld, & Laurent, 2015). Related to this, Eastwood, Frischen, Fenske, and Smilek (2012) attributes cause of aversive state to environment. According to them, boredom can occur when subjects are not able to direct attention towards internal or external stimuli sufficiently to engage in satisfying activity.

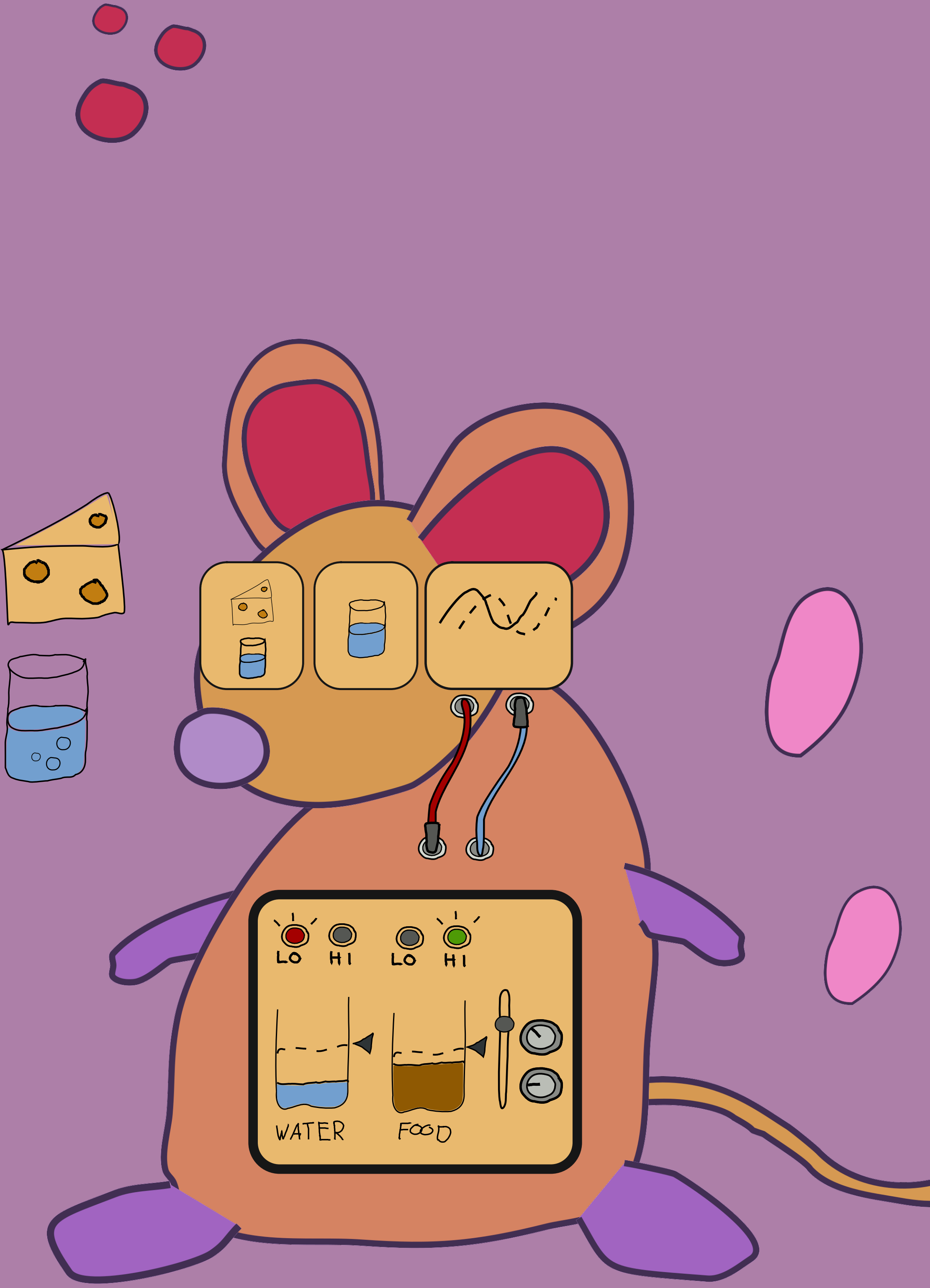
The results from these studies may be interpreted in terms of prediction errors. Specifically, boredom due to understimulation may be interpreted as insufficient prediction error, which motivates excitation–inducing novelty seeking, as described in the studies conducted by Bench and Lench (2019). In a complementary way, understimulation may motivate approach to uncertainty in the form of risk–seeking, but only if the person is able to appreciate the uncertainty involved (Yakobi & Danckert, 2021). In the absence of this appreciation, people appear to default to novelty–seeking. On the other hand, excessive prediction errors that cannot be coped with, may induce attentional lapses, a feeling of being overwhelmed, and depressive symptoms, as those described by Carriere et al. (2008). This corresponds to the apathetic boredom type proposed in Goetz et al. (2014). However, it is also possible that demanding tasks, i.e. tasks with high amount of novelty, but which are perceived as irrelevant or insufficiently motivated, can be met with the anger–like high arousal reactive boredom described by these authors.

In terms of calcium homeostasis, understimulation and a low rate of prediction errors should be associated with insufficient stimulation of NMDA receptors, and therefore an insufficiency of calcium inflow; a high rate of prediction errors would correspond to extended stimulation of NMDA receptors and an potentially excessive inflow of calcium.

4.5 Summary

In this chapter, I talk more about how allostasis is dependent on energy capacity and availability. Allostatic load can be understood as the energy cost necessary to rectify a physiological prediction error. For example, to increase the body temperature from 36 degrees to 37 degrees may involve activating muscles by shivering, and this shivering will consume some amount of energy. It will take more energy to increase temperature from 36 degrees to 37, than from 36.5 to 37, and the same typically holds for any kind of allostatic process. In the nervous system, diffuse signalling substances called neuromodulators complement more specific synaptic signaling. These neuromodulators may be connected to allostasis, and to the allostatic aspects of cognition. Serotonin and the serotonergic system may be understood as performing a sampling of environmental suitability, both in terms of protein availability and social support; dopamine may be understood as regulating approach to desired resources, including required nutrients, and is associated with psychological ”wanting”; the opioid system may be interpreted as measuring the energetic value of food, and to code for psychological ”liking”. Allostasis is perhaps most easily understood as physiological processes involving internal organs like the liver and the kidneys, but overt behaviour, like foraging and hunting, may also be viewed as allostatic. In this case, cognition may be seen as an integral part of allostasis since cognition may be understood as the control of behaviour in pursuit of allostatic goals. To elucidate this process, I follow allostasis in the nervous system from calcium homeostasis in single neurons, via regulation of excitation and inhibition at the network level, to behavioural reactions to novelty and boredom. The main point is that calcium concentration is tightly controlled inside neurons, and becomes toxic both at too low, and too high concentrations. Crucially, excitation and inhibition contribute to calcium homeostasis, since calcium is admitted by excitatory receptors in synapses. Given the predictive nature of cognition, and that prediction errors are sources of excitation, some optimal amount of prediction errors is necessary for neuronal health. This translates into optimal amounts of novelty, and insufficient novelty can lead to subjectively aversive feelings of boredom. Too much novelty, on the other hand, can give feelings of being overwhelmed.

Together these results can directly link information theoretic concepts like informational free energy, and philosophical concepts like predictive coding, to cellular physiological processes like calcium homeostasis and allostasis, as well as to psychological concepts like boredom.


LO
HI
LO
HI
WATER
FOOD

Chapter 5

Selection, attention, and choice

A moving organism in a noisy Umwelt must dynamically filter what information to use for choosing behaviour, and allostatic processes and their requirements are subject to prioritization. The environmental niche of an animal may be noisy, and full of signals that are not pertinent to that animal's needs. Being able to filter incoming information can significantly reduce energy demands on organisms, as can the ability to direct attention away from demanding stimuli. In the first section of this chapter, I explain when attention becomes necessary, and how it may be seen as a process that contributes to regulating computational demands. This keeps cognitive processes from being overwhelmed in busy environments.

5.1 Winner takes all - selection at the perceptual level

The central nervous system of most organisms are organized in such a way that signals, both from the outside and from inside the body, compete to influence overt behaviour. With regards to perception, an arrangement of mutual inhibition between populations representing different categories, ensures that a very limited number of percepts, or even just a single one, are allowed to influence motor behaviour. In this section, I go into some detail regarding the network motifs that support these processes.

Blakemore et al. (1970) presented human subjects with a visual display where tilted target lines were surrounded by varyingly different distractor lines. They found that, subjects tended to perceive the targets as tilted in the same direction as the distractors when the difference was small, but away from the distractors, when the difference was large.

Xiong and Chen (2002) cut olfactory bulb neuronal tissue from rats, and recorded electrical data from cells and dendrites. They report that inhibitory interneurons project GABA onto the dendrites of so–called mitral cells, that contribute to olfactory perception. Action potential propagation in mitral dendrites is thus regulated by inhibitory synapses that are distributed along the dendrites. These mechanisms contribute to *lateral inhibition* of olfactory perception, such that weaker olfactory signals are attenuated while stronger ones are propagated. Lateral inhibition thus entails that every neuron in the population can suppress every other.

For somatosensation, Baca et al. (2008) performed behavioral experiments on leeches while recording from neurons mediating responses to touch. They used intracellular microelectrodes to monitor the activity of several neurons at once, then knocking out inhibition both by pharmacologically means, and using electrophysiological techniques. The authors found that GABAergic inhibition among the motor neurons produced both lateral inhibition, as well as a generalized inhibition of most neurons in the central nervous system of the leeches. The generalized inhibition appeared to mediate the gain of the response, and enabled the response to different levels of sensory stimulation to have a broad dynamic range. In sum, the results show that very localized sensory stimulation of the leech's skin produces a balanced excitation and inhibition that sets the gain of the response.

These experiments show that at the network level, the winner–takes–all motif is mediated by means of lateral inhibition. The result of this arrangement is that only the few strongest neurons will propagate their signals forward in the system. Typically, at least in mammalian brains, the neurons carrying sensory information cannot themselves directly inhibit other neurons. Instead, inhibitive interneurons are excited by perceptual neurons, and project inhibitory GABA to the rest of the population.

5.2 Physiological needs affect attention by biasing salience

Mogg, Bradley, Hyare, and Lee (1998) presented words to human subjects, some of whom had a high level of hunger. The words were presented in two conditions; first, too quickly to be perceived and masked; and second, for long enough time that they could be adequately perceived. The hungry subjects showed a greater attentional bias towards words that were related to food, compared to subjects that were not hungry. The difference in attention was only present for words that were shown long enough to be perceived. According to the authors, these findings indicate that non–emotional motivational states like hunger can influence aspects of information processing in the brain. In particular, this is related to selective attention, and stimuli which are relevant to the current motivational state.

Using mice as subjects, Burgess et al. (2016) took a closer look at which brain areas might be involved in perceptual biasing for hunger. They used a microscopy technique called two–photon calcium imaging to image activity in primary visual cortex (V1), in postrhinal cortex (POR), and in feedback axons from the lateral amygdala to postrhinal cortex of the mice. The postrhinal cortex is part of the medial temporal areas that are involved with spatial processing, while the lateral amygdala is involved in conditioning (Erlich, Bush, & LeDoux, 2012). These authors reported that neurons both in POR and in lateral amygdala, showed increased activation in response to food cues while the mice were hungry. This was not seen in V1. The observed activity disappeared when the mice were sated.

Livneh et al. (2017) also investigated the effect of hunger on perceptual pathways of mice. They used a technique involving microprisms to image activity of neurons in insular cortex, while the mice performed a visual discrimination task. The insular cortex is in particular involved with processing internal signals from the body (Craig, 2002). According to them, insular cortex displayed strong activation when the mice were presented with already learned visual cues that predicted food rewards. Again, this activation disappeared after the mice had eaten, but appeared again if the researchers stimulated neurons in the hypothalamus that are related to hunger. Like in the study mentioned above, the results did not show much difference in V1 whether the mice were hungry or not. See also Burgess, Livneh, Ramesh, and Andermann (2018) for a review of how physiological needs can affect attention.

The experiments on mice indicates that early visual processing in V1 does not appear to be much affected by allostatic signals. To further elucidate this, Bogler, Bode, and Haynes (2011) looked into whether winner–takes–all (WTA) mechanisms could be distinguished from more graded forms of salience. Graded salience was defined by them to be related to varying contrasts between dark and bright areas of an image, while the WTA signal was defined to yield a single target. They showed natural scenes to human subjects in a fMRI scanner. Subjects were told to complete an effortful attentional fixation task, effectively directing their attention away from the natural scenes. The results indicated that signals in early visual cortex, and in posterior intraparietal sulcus (IPS) correlated with graded saliency. A WTA response, indicating the most salient position in the visual field, was in contrast associated with anterior IPS, and with frontal eye fields. In sum, graded and WTA saliency appear to be mediated by different neuronal populations. This agrees with the findings from the studies on mice.

As for the connection between WTA saliency and dopamine, Noudoost and Moore (2011a) investigated the response of neurons in frontal eye fields, and in V4 of occipital visual processing area of macaques. The frontal eye fields are held to be responsible for saccadic movements to targets in the visual field, while V4 has receptive fields for two– and three dimensional shapes, as well as for colours (Roe et al., 2012). In their experiments, the researchers altered activity mediated by dopamine receptors in the frontal eye fields by pharmacological interventions, and measured the effects of these interventions in area V4 neurons. They observed that the experimental manipulation was sufficient to change reliability and magnitude of V4 responses, but also the orientation selectivity of the neurons. According to the researchers, the changes in the V4 neurons were comparable to that of top–down attentional signals. These results thus shows how dopamine is involved in WTA selection in area V4 via modulation of neurons in the frontal eye fields. See Noudoost and Moore (2011b) for a review of how neuromodulators affect selective attention.

Dopamine projections to prefrontal cortex, including the frontal eye fields, originate in the ventral tegmental area (VTA). B. L. Lewis and O'Donnell (2000) tested whether stimulation of this area could change electrical states of cortical pyramidal neurons in live rats. They inserted stimulating electrodes into the rats' VTA, as well as their thalamus and hippocampal areas, and recording electrodes in the rats' prefrontal cortices. According to them, prefrontal pyramidal neurons alternate between a down state with negative cell potentials, and an up state where the cells emit action potentials. Results indicated that stimulation of thalamus and hippocampal areas were not sufficient to drive the prefrontal cells into the up state. Stimulating the VTA appeared to give rise to a variety of spiking behaviours. In particular, burst–like stimulation induced comparatively long–lasting up states. These behaviours could be blocked by dopamine antagonists. Chemical stimulation of the VTA also produced long–lasting up states. Together, these results indicate that dopaminergic projections from VTA contribute to regulation of activity in the rats' prefrontal cortices.

Moving backwards from the VTA to the hypotahalamus, Nieh et al. (2016) used a combination of chemical– and optical stimulation techniques to elucidate how the lateral hypothalamus affects the VTA in live mice. The lateral hypothalamus projects both inhibitory GABA neurotransmitters, as well as excitatory glutamate to the VTA. The results from this study indicates that lateral hypothalamus has a disinhibitive effect on dopaminergic VTA neurons, such that they selectively inhibit local GABAergic interneurons in VTA. This effectively increases dopamine release into the nucleus accumbens of the basal ganglia. Behaviourally, stimulation of the inhibitory part of the pathway from lateral hypothalamus to VTA, appeared to support positive reinforcement and place preference in the mice. In contrast, stimulation of the excitatory glutamate part of the pathway appeared to mediate avoidance of places. According to the researchers, their results are consistent with the lateral hypothalamus and the VTA playing roles in increasing motivational salience.

Taken together, the studies mentioned above outline pathways that contribute to finding food, and maintaining allostasis. An interpretation of these results in the context of allostasis, sketches a picture where signals from the body e.g. about energy levels, can influence the hypothalamus to selectively increase dopaminergic projections to frontal eye fields. In turn, the frontal eye fields can modulate visual areas like V4, such that food cues become more salient for an animal. Associations between subcortical populations involved with allostasis, and the dopaminergic system can therefore regulate the perceptual system, biasing it towards required resources.

5.3 Attention limits computational demands of cognition

Some perceptual biases are related to properties of the percepts, such as contrast. Various kinds of grouping properties, like breach of continuation, may also bias attention. In addition, attention may be influenced by current allostatic prediction errors, like low blood sugar. In such cases, groups of percepts may be modulated such that they are more likely to become active, yielding strong activation. In effect, percepts thus become salient: they draw attention and stand out. In any case, the result of biasing and selecting percepts, is that information going into upstream processes is limited to stay within the computational bounds of those upstream processes.

Visual attention is closely associated with eye movement and saccades, as we saw above (Noudoost & Moore, 2011b). In effect, saccades turn perception into a sequential process, sampling areas that are marked by the winner–takes–all process described above as containing strong signals, and therefore as having potential interest. This process massively reduces the processing demands of upstream populations, ensuring that metabolic demands are not exceeded, and that the animal maintains its grip on its environment (Bruineberg & Rietveld, 2014). In practice, it means that populations responsible for choice and behaviour can act quickly enough to keep the animal alive. Although information sampling is a sequential process, there is still the subjective impression that the world is stable and whole. I will go into two mechanisms that contribute to this impression: spatial indexing, and synchronization of the sampling systems with episodic memory systems.

In Balkenius et al. (2018) we present a memory model that included a network for keeping track of locations in the environment. This made use of the winner–takes–all motif explained above, such that as the attentional window samples the environment, only one location can be active at a time. This mechanism can work in both a bottom–up as well as a top–down fashion. Hence gaze can follow a path of interest determined by the environment as described above, but it can also be biased to go to locations that are determined by higher order, goal oriented systems.

Fiebelkorn and Kastner (2019) theorizes that this kind of spatial indexing, or spatial attention is synchronized with episodic memory areas, including the hippocampus by means of the theta frequency band. Both saccades and the hippocampal system operate at this frequency, allowing them to be synchronized. Johansson, Holsanova, and Holmqvist (2006) showed that listening to verbal descriptions are associated with eyes saccading to different spatial indices, as a mental image is being built up. This indicates a tight association between spatial attention and eye movement.

5.4 Choosing what to do, and choosing where to go

According to Passingham and Wise (2012), behavioural choice in mammals is mediated by medial prefrontal cortical (MFC) areas. The MFC, excluding the nearby orbitofrontal cortex (OFC), is connected in such a way that it is closely involved with foraging and allostasis. Firstly, it is connected to the hypothalamus, which is responsible for processing prediction errors from the body. Secondly it is connected to the anterior cingulate cortex, which processes the valence of these prediction errors. Thirdly, it is connected to the hippocampus which is involved with coding places that contain resources and means for reducing allostatic prediction errors (Kennedy & Shapiro, 2009). This connectivity also contributes to the selection of approach strategy in foraging situations, that is whether to engage in direct approach, or detour behaviour (Patai & Spiers, 2021).

In Tjøstheim, Johansson, and Balkenius (2021) we simulated situations of such strategy switching with a model featuring a simplified pathway involving MFC dynamics. This model indicates that MFC likely contributes in an inhibitory way, suppressing automatic direct approach in the face of obstacles that require detours.

After having navigated to a foraging site that contains needed resources, the OFC is engaged in choosing the objects that are most likely to reduce allostatic prediction errors (Passingham & Wise, 2012). Related to this, the OFC has connections to areas that are involved with perception of food items, including the dominant visual areas, but also olfaction, gustation, and visceral areas. The OFC also connects to the amygdala such that e.g. ingestion outcomes can be fed back and update predictions. While many parts of the OFC are involved with specific outcomes, some parts are concerned with more general outcomes, by Passingham and Wise (2012) referred to as a ”common currency” valence. Referring to the above chapter on energy and allostasis, an intriguing possibility is that this common currency is in fact related to metabolic energy. That is, energy gain or energy saving.

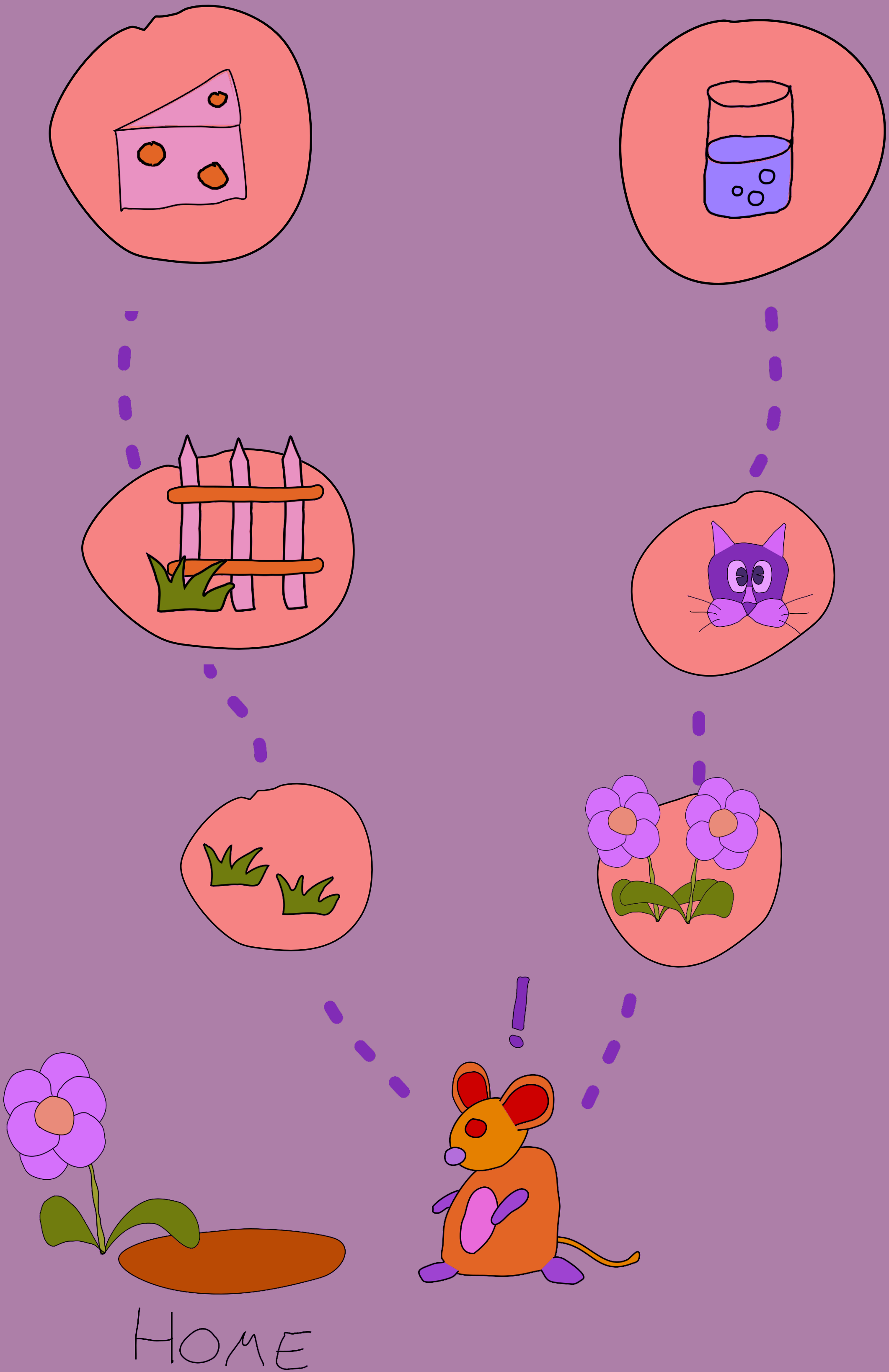
The notions of foraging described above may appear abstract and removed from everyday human concerns. But as Gidlöf, Anikin, Lingonblad, and Wallin (2017), and

Gidlöf, Wallin, Dewhurst, and Holmqvist (2013) show, they can be related to commonplace human activities such as shopping for food in the supermarket. In these studies, people are fitted with portable eye trackers that record eye movements. Gaze patterns are mapped to images of supermarket shelves, thus enabling analysis of attentional paths in a real life situation. One salient result from these studies is that looking time tend to predict choice and purchase behaviour.

We used aspects of this in Balkenius, Tjøstheim, et al. (2020) presenting biologically plausible network motifs that can support this kind of temporal accumulation of value. We combined this with the memory model described in Balkenius et al. (2018), that provides support for recognition and spatial indexing, as well as episodic–like memory. Hence this model shows how the object choice process could work in the brain, and that simulation of future value can contribute to evaluation of objects, when choice is not biased by immediate allostatic needs.

5.5 Summary

In this chapter, I have introduced the notion of perceptual filtering based on signal strength and salience. Since the environment is full of change and noise, organisms have evolved to only take in as much as is required for allostasis. Winner–takes–all mechanisms in the perceptual networks assure that only the strongest signal is propagated through the system. As was discussed in the previous chapter, these strong signals are typically also prediction errors. In this way unexpected changes in the perceptual field attract attention. Spatial aspects of perception and attention, where immediate surroundings are coded as a kind of sensory globe, ensure that eye focus, head direction, and body orientation can be coordinated to make up ”places”. A place may offer up several opportunities, and an opportunity may be taken advantage of in alternative ways. Mammalian brains have particular networks that appear to have evolved to cope with options and choice, such that the orbitofrontal cortex biases choice of objects based on experienced outcome value, like taste; networks in the medial prefrontal cortex appear to be involved in the choice of behaviour to get at those objects.


!
HOME

Chapter 6

Foraging, navigation, and the navigation–place metaphor

Control and regulation of foraging behaviour are fundamental responsibilities of cognition. This implies directing the organism in space, and hence *navigation* tend to be a primary concern for many cognitive systems.

In this chapter, I zoom out to the level of the foraging animal, and discuss navigation processes in the context of allostasis. I then zoom in to how signals from the body, in the form of allostatic prediction errors, drive the eliciting of foraging. Effective navigation between an environment's resource sites increases demands on learning. The next section explores this, which then leads into the notion of planning to reduce foraging costs. I then attempt to show how the machinery involved in navigation may be generalized to predict complex situational–dependent behaviour that is different from navigation.

6.1 Control of foraging behaviour is a fundamental responsibility of cognition

Keeping with the starting point of cognition as recognition and choice, spatial cognition might be understood as recognition of location or place, and the choice of which path to take e.g. to get from a nest to a foraging site. Expanding to more elemental processes, the perception of place, and where in space to direct attention, are also part of spatial cognition.

In this section, I will consider primarily navigational space, and leave manipulation space, the part of space that can be reached by manipulators, to the next chapter. Navigational space is of a dimension that typically requires travelling, and may be thought of as consisting of more than a single *place*. Depending on anatomical affordances, travelling may be done by organisms by belly, fin, foot, or flight.

Places are important enough that many species have neuronal populations that are specifically tuned to process them (Nakazawa, McHugh, Wilson, & Tonegawa, 2004; O'Keefe, 1976). Place cells are typically found in mammalian brains in areas medial to the temporal cortex, and most prominently in the hippocampal formation. In foraging, place cells may be used both to give indications of distance to a goal–site, but also to recognize the goal–site itself (Xiao, Lin, & Fellous, 2020). Topologically, place cells appear to be associated with a kind of sensory globe, the place field, which itself is represented in sensory association areas in the lateral intraparietal cortex (Kesner & Creem-Regehr, 2013). Hence when an animal turns its head, the place field remains constant even if the sensory input changes (Weber & Sprekeler, 2018). In humans, Johansson et al. (2006) showed that eye movements reflect positions of objects described by spoken word, either heard, retold, or redescribed; this effect was present both while watching a blank screen and in total darkness. Thus it appears that eye movements may also be associated with spatial indices in the sensory globe, and may hint at a way that head direction cells and saccades together contribute to place fields. Since the sensory globe is based on the animals point of view, it refers to *egocentric* space. This is complemented by a third–person perspective called *allocentric* space.

Although medial temporal areas contain many specialized populations that take care of e.g. environmental boundaries and obstacles (Poulter, Hartley, & Lever, 2018), so–called grid cells are of special interest (Hafting, Fyhn, Molden, Moser, & Moser, 2005). These cells keep track of allocentric space, which is map–like and independent of place. Grid cells are connected such that they form a kind of doughnut (Gardner et al., 2022). This means that as the animal moves, grid cells representing the surface of this doughnut will be activated. There are indications that different populations represents doughnuts of varying sizes, allowing the animal to keep track of where it is on a range of spatial scales (Bush, Barry, Manson, & Burgess, 2015). Together, place cells and grid cells mediate recognition of environmental features both in turns of what a place "looks like", as well as where the current place is in relation to other places. This latter feature can then be used by the animal to e.g. find shortcuts (Banino et al., 2018) or alternate routes; but also to direct itself to neighboring places by means of spatial inference. However, when an animal arrives at a foraging site, and comes within view of a resource, it will often be compelled to approach it directly, even if something is in the way. Approaching resources thus may involve choice between charging ahead or making a detour.

In Tjøstheim et al. (2021), we presented a neuronal spiking model of a network that could use both direct approach to a sighted resource, as well as use map–like processes to take detours and move around obstacles. This model used a simplified sensorimotor pathway, based on the mammalian spatial processing pathway, as well as a black box representation of grid cell populations to support detours by using a map. The model controlled an agent in a virtual 3D environment, and the network was designed such that direct approach is a kind of bottom up process, activated by perception of a resource. Detour behaviour on the other hand, requires inhibition of the automatic approach pathway. Our results indicates that the amount of inhibition an organism can exert to stop direct approach, will affect how successful the organism is in getting around obstacles. But the results also indicates that the size of the inhibitive populations can play a part, by reducing strategy variation. That is, with small inhibitive populations, the organism may revert back to attempting direct approach before a detour is completed, reducing its success rate.

To sum up, one of the core responsibilities of a cognitive system is to support an animal when moving through space to find food or other resources. This is important because different resources tend to be found at different places. There may also be more than one path to the same place, each with their own benefits and drawbacks. Recognition is thus used by animals to understand where they are, and where to head next, while choice circuits are necessary to allow them to use the most beneficial path to get to a goal.

6.2 Allostatic prediction errors drive foraging

It may sound contrived to describe an animal as ”predicting” that the level of blood glucose is at some level, or that the level of sodium is another level, and that they are ”surprised” when those levels are not what they should. However, these predictions must be understood in the context of a control system which must continually anticipate changes and adjust setpoints accordingly. Although it is likely that many animals have some subjective awareness of these differences, and that this awareness is similar to human hunger and thirst, we can still understand much of what is going on without invoking subjective consciousness.

The pathways in mammalian brains that ultimately drive foraging are relatively well known. According to e.g. Shadlen and Kandel (2021a), the hypothalamus coordinates allostatic processes, and receives information about the state of the body from the peripheral nervous system. Hypothalamus projects to cingulate cortex via thalamus. Cingulate areas appear to both do interoception of signals from the body, as well as processing the valence of interocepts (Craig, 2002). That is, whether they are

”good” and may be left alone, or ”bad” and hence in need of amending (J. Gray & McNaughton, 2000). Cingulate cortex in turn has rich connectivity to motivational networks in the basal ganglia that can initiate behaviours.

In this chapter, the focus is on navigation and places, and hence it is not only interesting how movement is initiated, but also how the brain determines which place to mark as a goal. Kleckner et al. (2017) provide evidence for a network that is responsible for interoception and allostasis, that includes both subcortical structures like the hypothalamus and amygdala, as well as cortical areas like cingulate cortex. They provide also evidence that this allostatic network interacts with the default network and the salience network, both of which are primarily cortical. One interpretation of this is that the allostatic network biases the default network to construct ”images” in working memory of needed resources, while the salience network is biased to increase the saliency of signs that predict those resources, as well as instances of the resources themselves. Hence the feeling of hunger is accompanied by images in working memory of food, as well as simulations of eating. At the same time, motivation to approach food sources increases. Crucially, the spatial networks can then be tasked with coming up with concrete goal–sites where the required resource may be found, as well as possible navigation paths to those sites from the current position of the animal.

Related to this, Chanes and Barrett (2016) theorize that there is a flow of predictions and prediction errors along cortical laminar gradients. That is predictions flow along a gradient of increasing laminar differentiation. This pathway goes from the limbic cortices, that has less laminar differentiation, to primary exteroceptive cortices, like visual and auditory areas that have typically six clear layers and with a pronounced layer 4. Prediction errors then flow in the opposite direction, carrying the difference between what was predicted and what sensors report. The authors refer to limbic cortices as including the anterior- and middle cingulate cortex, as well as the ventral anterior insula. Hence prediction errors are initially detailed and modal, but become increasingly compressed and integrated as they come closer to limbic areas. Importantly, the converse is the case about predictions; they are quite abstract at initiation, and hence can be realized in many concrete ways. This means that potentially many types of food can be recognized to reduce interoceptic prediction errors, and may be found at any opportune place.

Even if an animal is neither hungry nor thirsty, it may still be motivated to explore. One way this motivation can develop is through the energy dissipation of the brain itself. Recall that informational prediction errors carry metabolic cost (Peters et al., 2017). This can make novelty aversive to an animal and reduce motivation to explore, since excessive energy demands tend to induce aversion (Dunn & Everitt, 1988).

In chapter four, we looked at regulation processes in neurons, and at calcium regulation in particular. Neurons must be stimulated to avoid becoming hypoactive. This could mean that at the behavioural level, under–stimulated populations may motivate approach to stimulating prediction errors in the form of novelty, and novelty–seeking may hence be regarded as allostatic behaviour

A complimentary way that information seeking may be motivated is by means of the reduction of uncertainty. Uncertainty is metabolically costly (Peters et al., 2017) perhaps because it means keeping neuronal populations active that represent several possibilities, or outcomes (Geurts et al., 2021). This kind of "epistemic foraging" may not be directed towards novelty as such, but rather the elimination of possibilities (Mirza, Adams, Mathys, & Friston, 2018), and the attenuation of excessive neuronal activation.

In all these cases, learning is involved; the animal must learn where foraging sites are, and where to go to experience novelty. Next we will look at how this learning process works.

6.3 Effective foraging is supported by learning

One of the functions of the dopaminergic system is to contribute to the learning of signs, or cues, that predict rewards. Some evidence for this from experiments on rats, is that dopamine lesions in nucleus accumbens prevented sign–tracking but not goal–tracking (Chow, Nickell, Darna, & Beckmann, 2016). Here sign–tracking was operationalized as pressing a lever in response to a tone, while goal–tracking was counted as direct approach to food. Colaizzi et al. (2020) mapped this experiment onto human behaviour and showed that sign–tracking behaviours in rats appeared to overlap with addictive behaviours in humans. The significance of this in a foraging context could be that signs, like the sound or smell of a prey, contributes to motivation to approach, even if the prey is not directly observed. The dopaminergic system thus contributes by making sure that signs can be learned and recognized. Next we shall see how neural populations in the medial temporal area may keep track of such signs.

According to Poulter, Lee, Dachtler, Wills, and Lever (2021), vector trace cells (VTC) in the subiculum, appear to mediate spatial relationships between an agent and multiple cues. VTCs thus embody a global integrated representation of the entire environment. The subiculum apparently incorporates representations of both environmental boundaries, as well as objects and their relationships. Overall, VTCs suggests a vector based model of computing spatial relationships, and one that is freed from constraints of direct perception of cues. This enables spatial planning and imaginative cognition. The former because it affords an organism options of alternative routes; the latter because abstract cues may be concretized based on experience. In this way, imagination can be used to play out different scenarios which can in turn contribute to planning. Choices need to be made though, and next we will look into how optimal foraging theory can contribute to this.

Papini (2002) presents an ecological view of learning where learned behaviour is part of the organism's biological equipment, allowing for an adaptive fit to its environment. According to optimal foraging theory (Charnov, 1976), natural selection shapes decision rules so that behaviour maximizes resource value, and minimizes behavioural costs. When an organism is at a foraging site, there are several aspects that a cognitive system typically takes into account. The first is the reward of the patch the animal is at; the second is the learned average rate of the reward of the environment, including alternative sites; the third is how much it would cost for the animal to leave the current patch and find another (Wallis & Rushworth, 2014). As detailed in Wallis and Rushworth (2014), for mammals, the anterior cingulate cortex learns the value of switching to a new patch, while the ventromedial prefrontal cortex is involved in making the choice of where to go, once the decision to switch has been made.

In Balkenius, Tjøstheim, et al. (2020) we presented a model that employed the use of different memory mechanisms to support decision making. In particular, we modeled making decisions when the alternatives could not be evaluated based solely on immediate sensory information. Instead, the model employed a type of prospection, or "imagination", to evaluate a possible future that could result from choosing one of the alternatives. We used the everyday process of selecting groceries in a supermarket as a running example, but the same process could plausibly be used by foraging animals. In fact, Redish (2016) reviewed how mice apparently "play forward" sensory impressions associated with taking different routes at a choice point in a maze. Experiments indicate that this playing forward process is mediated by sequentially activating place cells associated with each of the two paths. The vicarious trial and error process further appeared in particular at points where the cost of making a decision is high. That is, when the choice is not obvious, and needs some deliberation to work out which is the better option.

6.3.1 Sequence memory and object permanence supports way–finding

The choice–model presented in Balkenius, Tjøstheim, et al. (2020) is based on a model of memory proposed in Balkenius et al. (2018). This model was mentioned above in the context of recognition processes. Again, the three central components are an identification network, a localization network, and a working memory network. In the model, attention serves as the interface between the inner and the external world, directing information from sensory organs to memory. Also, attention controls top–

down influences on perception, thus mediating sensory expectations. Of interest to this section is the sequence memory which may be employed to learn paths between forage sites containing different resources. Also of interest is the working memory, since it potentially supports maintaining the image of way points. The usage of way points is part of an allocentric navigation strategy that relies on learning of direct sensory observation of salient features, along a path. Examples of such way points may be environmental features such as patches of brightly coloured plants, or discontinuities of vegetation. Hence way–points may be considered as a form of sub–goals. Shamash et al. (2021) showed how mice used the strategy of memorizing way points when making their way back to their nest in the face of threat, but also in a similarly structured food–seeking task.

The ability to learn a sequence of sub–goals and the availability of a working memory affords animals the ability to partition a complex path to a foraging site into segments. The recognition of one way–point can then trigger the next one to be held in working memory. The encounter of a way–point likely also works as a kind of sign that progress is being made (Carver, 2004, 2006), which then should periodically trigger motivation to continue approach. In a complex environment, way–points may also serve as reassurance for the animal that it is on the right path (Caduff & Timpf, 2007).

6.4 Machinery for navigation lends itself to more generalized prediction: the navigation–place metaphor

Above we heard about the process of prospection, or vicarious trial and error apparently performed by mice at choice points (Redish, 2016). According to Barsalou (2009), *conceptual processes* may be regarded as special cases of simulation processes. He argues that concepts are most often bound to a situational context, and also to some specific goal that a person is approaching, or trying to achieve. In his example, a bicycle may be conceptualized to fill several roles depending on the goal. For one person, the bicycle might be a means of transportation to get to the supermarket. For another, a bicycle is something to be repaired in exchange for money to buy food, and pay the rent. In this interpretation, abstract concepts can become quickly meaningless in the sense that they lose their behavioural associations. Abstract concepts thus need to be made concrete, and related to behaviour, and some end goal, to become meaningful.

Consider the following somewhat idiosyncratic example: Lisa, a member of a musical group wants to play a concert at a club, and is being tasked with making it happen. In her case, the goal is not a place as such, but rather ”to have successfully played a concert”. This goal involves many sub–goals each with complex situational–dependent dynamics. Let's go through a few of them. First, Lisa drives to the club to check out the venue and learn what the stage looks like, and what kind of equipment the club can provide. Some weeks before the concert, Lisa notices that her guitar is broken. She takes the guitar to a repair shop to fix it. Fixing the guitar is expensive, and Lisa is out of money. She promises to pay for the repair with some of the money she will get for the concert. As collateral, she gives the repair shop her only golden necklace.

Translating this into processes involving spatial cognition and situational conceptualization, we might get something like the following. Since Lisa has experience with playing concerts in other clubs, she can visualize many potential hangups if the band shows up unprepared. Hence Lisa chooses to both learn the way to the club, as well as talk to them about equipment, and inspect what the stage looks like. When she notices the broken guitar, she understands that without the guitar, she won't be able to play at the concert. This generates a new sub–goal of returning the guitar to working order, where the guitar also changes its role from being something that makes sound, to being something that requires effortful and expensive maintenance. At the repair shop, and when hearing the price of the repair, Lisa changes the conceptual role of her necklace from jewelry to collateral. A complex prediction ensues as she hands over the necklace. It involves her getting the repaired guitar, playing the concert, getting paid, going back to repair shop and paying them, and finally being handed back her necklace. That is, a string of navigation between goal–sites, and accompanying situation–dependent behaviour. In the process of all of this, the club has also changed conceptual roles several times, from being a site of epistemic foraging, to a site of performance and intrinsic reward, to a site of extrinsic monetary reward.

In Tjøstheim and Stephens (2021) we proposed an interpretation of "intelligence" as accurate prediction, and of "general intelligence" as decontextualized accurate prediction. This may be synthesized with Barsalou (2009)'s notion of goal–directed conceptualization. In particular, simulations and predictions may be seen as related, such that predictions are simulations of real–world processes that can influence choice and behaviour. But simulations may also be used to construct fictions that have no bearing on future behaviour. Hence Barsalou's concepts may be interpreted as simulation–models that somehow afford playing out of diverse scenarios. The accuracy of these scenarios when employed as predictions depend on experience, and the difference between predictions and outcome are naturally coded as prediction errors. Hence being able to learn from these errors can increase accuracy over time. In the context of intelligence, there is also an aspect of how long into the future a prediction stays accurate. Short predictions have a much higher probability of being correct than do long ones, due to continually branching possibilities, and the inherent noise of reality. However, these branching possibilities may be constrained by knowledge, which may contribute to motivation for epistemic foraging (Gottlieb et al., 2013).

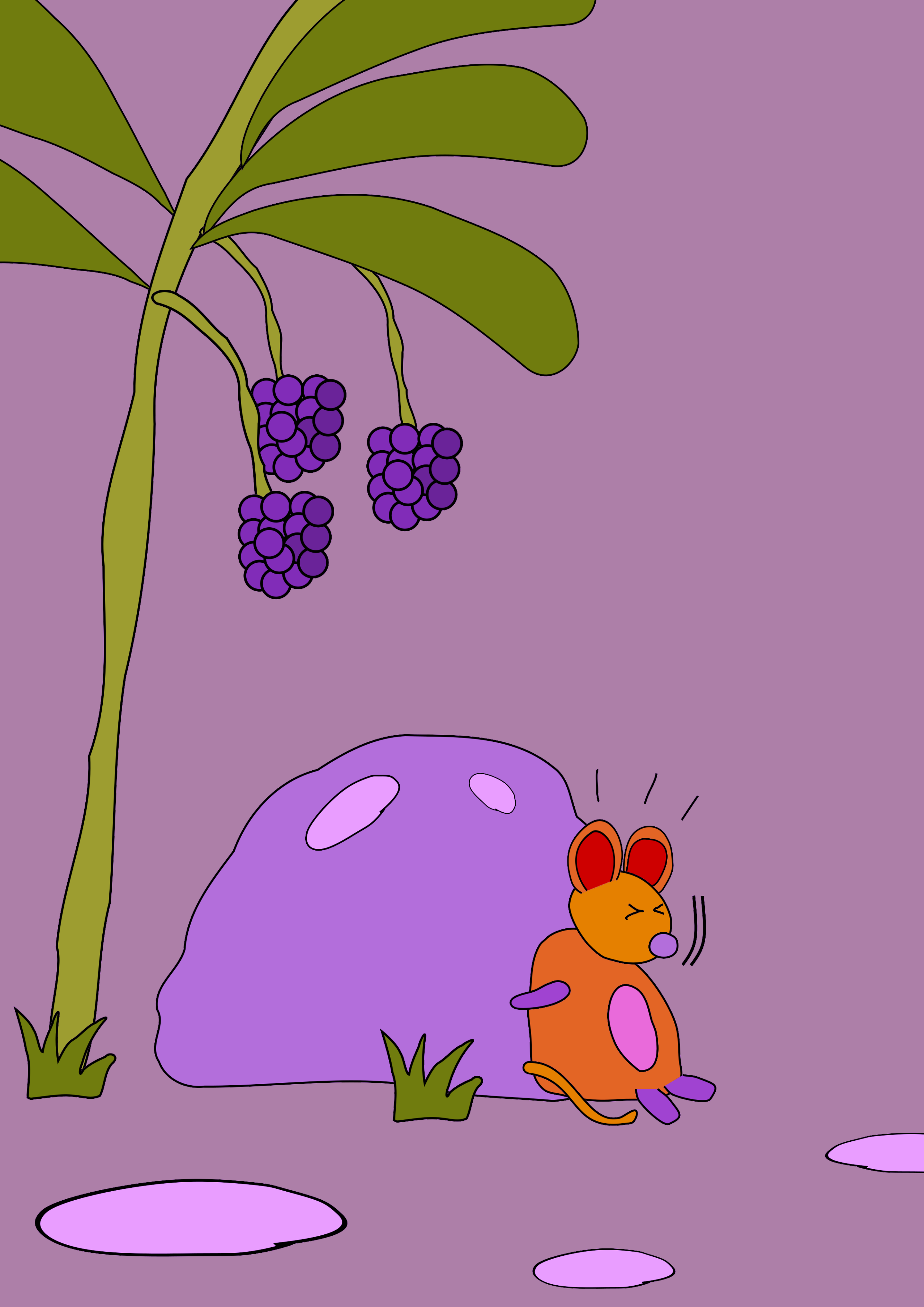
In a purely spatial sense, knowing all the possible paths will help constrain simulations of which path is the fastest, or the least energy costly. It also allows an animal to take detours, and use alternate paths if the quickest path is somehow unavailable. Mammalian brains have particular machinery for representing spatial boundaries and borders Doeller, King, and Burgess (2008); Poulter et al. (2018), and these representations are powerful means to constrain simulations, and make them computationally tractable. Spatial boundaries may then perhaps be abstracted into more general notions of limits, including ”least” and ”most” (Woodin & Winter, 2018).

Navigation has also strong embodied aspects that involve the movement of the animal, and how much that movement costs in terms of energy (Rushworth, 2008). Joints have constraints too, as do cardiovascular capabilities. These embodied associations contribute to simulations of navigation along different pathways by tagging them with costs, and tractability. Mental behaviour is analogous to physical behaviour in that it is energetically costly, and subject to individual physiological limitations (Gevins & Smith, 2000), and should thus also be amenable to simulation.

Hence by employing the navigation–place metaphor, humans and other mammals can make accurate predictions in domains that are not spatial, but which nevertheless have some spatial properties. Fortunately this is the case of many processes, including human projects. But for complex projects, just navigating between way points is not sufficient. There are also tasks to accomplish at each way point. In the next chapter, we will consider the manipulation–state metaphor to fill in this requirement.

6.5 Summary

In this chapter, I have discussed how navigation in the context of foraging is one of the main responsibilities of cognition. Using the ”place” percepts presented in the previous chapter, the medial temporal areas of the mammalian brain can encode foraging sites, nest sites, and paths between them. This allows an animal to navigate between a safe place where it can sleep, and places that have opportunities for finding food or other resources. Foraging behaviour is motivated by allostatic prediction errors, coupling the internal physiological environment of an organism to the external environment. This coupling is made more specific by the modulation of perceptive salience by dopaminergic projections, and possibly by forming ”goal–images” in working memory. As we saw earlier, sensory stimulation may be a motivation too, and exploration may thus be viewed as a kind of foraging for novelty. In a variation of this, uncertainty about states of the world can give aversively high stimulation, but can be reduced by foraging for knowledge, or epistemic foraging. This is only possible if the cognitive system supports learning, or adaptation that can reduce prediction errors. In particular, learning of sequences is necessary to be able to learn the path to a foraging site that is not immediate observable from the nest site. Finally, machinery for spatial cognition may be applied to predict even abstract sequences, but that have spatial–like properties. In particular, medial temporal networks that use a grid–like coding affords many kinds of relational inference, including transitive inference. In spatial terms, this means that if I can go from my home to the bus stop, and from the bus stop to my job, then I can go from my home to my job. This general pattern is appropriated in abstract logic, and thus affords more general reasoning abilities.



Chapter 7

Obstacle negotiation, manipulation, and the manipulation–state metaphor

A responsibility of a cognitive system is to maintain approach to foraging–sites in the face of obstacles. Obstacles may be negotiated by means of detours, as discussed in the previous chapter, but also by means of manipulation.

Environments may pose obstacles to a foraging animal both in the form of geography, and incidental environmental configurations, as well as predators. In this chapter, I first make an account of how niche variability may drive organisms to become better at generalization, then attempt to explain the distinction between variability and complexity, and how complexity can drive abilities to predict. Next I make a connection between niche complexity and the energy cost associated with making accurate predictions. The ability to manipulate environments may help energy extraction in complex, obstacle rich environments. I look at the processes involved in manipulation next, and how effective manipulation is supported by the learning of skills. Similar to how navigation processes may support prediction, manipulation abilities can support the use of tools, and the construction of technology. In the final section in this chapter, I explore these aspects of cognition.

7.1 Generalization requirements grow with niche variability and food availability

Bruineberg and Rietveld (2014) propose that an ”econiche” can be understood as a ”landscape of affordances” for an animal, and is dependent on the animal’s body and information processing machinery. In Tjøstheim and Stephens (2021), we argue that such niches may be usefully considered as more or less variable . The number of different species in an animal’s diet is one way of operationalizing this variability, as is the number of habitat zones, such as dense forest, grassland, or water. The former affects how an animal recognizes food, the latter affects how it must be able to move to navigate to food sources. A panda is a specialist, recognizing only bamboos as food. A rat is a generalist, and will eat grain as well as meat. Econiches hence may allow a species to be particular about food via preferences, but could also force them to be particular because of limited options. In either case, changes in availability of preferred food, or of the single available food, can impose on animals the need to relax their ”definition” of food, and hence to generalize.

Food can be considered a perceptual complex, and are made up of associations of many sensory modalities including smell and vision. When generalization happens, it can entail that detail information in the receptive fields of neurons become blurred, but also that the number of neuronal associations grow. In a practical sense, the number of sensory impressions that trigger approach and consummatory behaviour increases in the foraging animal.

Variability is related to complexity, but in the next section we will look at complexity from the perspective of prediction, and how the complexity of a niche can pressure species to become better at prediction.

7.2 Predictive requirements grow with niche complexity

In Tjøstheim and Stephens (2021), we argue that cognitive–behavioural arms races can develop between predator and prey species, such that prey are pressured to behave in ways that are increasingly unpredictable, while predators must compensate by developing their predictive capabilities. In that paper we distinguish between ”adverse” and ”nonadverse” obstacles to resources, where adverse obstacles are typically organisms that have cognitive capabilities and can control their behaviour, and hence actively compete for those resources. Nonadverse obstacles do not display such capabilities, and are typically exemplified by hills, rocks, and rivers.

The presence of cognitively capable predators and adversaries increases the complexity of a niche in ways that are qualitatively different from the variability discussed in the previous section. Another way of putting this, is that while variability can pressure species to become more cognitively flexible, adverse obstacles can pressure them to become more cognitively *sophisticated*. This notion of sophistication may then be interpreted as both predictive abilities, but also complimentary abilities to plan behaviour to take into account predictions. Sophistication typically comes with costs though, which we will look at next.

7.3 Prediction becomes increasingly energy costly as niche complexity grows

Geary (2018) argues that mitochondrial capacity may be an important factor in what determines individual differences in the general intelligence measure called *g*. Energy is involved in cellular processes, as we have seen in earlier chapters. Geary mentions in particular aspects such as the number and capacity of mitochondria in presynaptic buttons. These make a difference in a button's capacity to both produce and recycle neurotransmitters, and hence can make a difference in cognitive processes such as working memory maintenance.

Generally, the number of neurons in a brain, and the number of connections, make a difference to how much detail that brain can process (Kolb, Pedersen, Ballermann, Gibb, & Whishaw, 1999; Silva et al., 2021). Accurate timing is also necessary, but is also costly in terms of needing both myelin–rich connections, as well as various kinds of high frequency interneurons. The latter are used e.g. to produce precise neural trains of action potentials by inhibiting short subsets of spikes, or even single spikes via so called sculpting inhibition (Shadlen & Kandel, 2021b).

Recent research indicate also that the cerebellum may be involved in complex cognitive function (Mcafee, Liu, Sillitoe, & Heck, 2022). This appears to be by means of precise manipulation of the phase of neuronal frequency bands, which regulates communication between different cortical areas.

Given that the neural machinery for sophisticated predication and behavioural control exists, we will next look at how manipulation in particular can afford access to resources.

7.4 Manipulation affords extraction of resources from complex, obstacle rich niches

Changizi (2010) describe the ”leafy branch habitat” that may have contributed to several useful evolutionary adaptations in the lineage of humans and apes. According to them, this habitat may have pressured hominid descendants to use fore–limbs for *grasping* in addition to walking. Combined with improvements in motor control, visual acuity, and color vision, this adaptation opens up a range of new affordances for *manipulation*.

Manipulations affords reconfiguration of the local environment, such that otherwise unavailable resources become available. Such reconfiguration may entail removal of obstacles, as well as generally moving things around to bring food close enough to eat. Experiments on corvids and apes contain a wealth of examples of manipulation behaviour; much of these behaviours apparently imply sophisticated simulation of physical processes and mechanics in particular. Bird and Emery (2009) showed that corvids can use pebbles to bring a floating reward in a glass into range, by dropping the pebbles into the glass. Similarly, apes could move boxes and stack them such that they could be used as a platform to reach rewards (Köhler, 1925). Corvids have also been observed to drop nuts onto hard surfaces, using them as nutcrackers (Zach, 1979). Manipulations can become intricate and require precise motor control. This can be supported learning skills, as we shall see next.

7.5 Effective manipulation is supported by learning skills

Where sensory information starts out detailed and get progressively more compressed, behavioural and motor information do the opposite. They tend to start out, in the case of manipulation, with an overall goal in the form of a desired state of the world. This goal is then made concrete through the motor pathways, until motor fibres are stimulated, and movement is produced. Hence skill is supported by the parietal ”how” perceptual pathway, that provides spatial feedback information about where both goals and manipulators are positioned relative to each other (Milner & Goodale, 2008).

Skills can also be viewed as ”knowing how”, which involves precisely coordinating muscle contraction and relaxation, but also includes being observant of the process of state change (Beilock, Bertenthal, Mccoy, & Carr, 2004; Williams & Davids, 1998). This means that the execution of skilled movement may be viewed as a continuous regulatory process, where motor movement is employed to adjust prediction errors, as the manipulation goal is brought into being. According to Ito (2013), the cerebellar microcircuit learns both to predict sensory signals, as well as to produce control signals that keep sensory signals within some bounds. Hence the cerebellum contributes to skill learning by gradually making movements more precise, but also increasingly independent of conscious and effortful control.

Manipulation involving tools can expand affordances, but tools may need to be carried on foraging trips. This can increase the need for planning as we'll go into next.

7.6 Tool use can increase need for planning

Incidentally found objects may be used in a tool–like way, as when apes use sticks to capture ants (van Lawick-Goodall, 1971), or stones to crack nuts (Visalberghi, 1990). Humans routinely carry specialized tools with them, to carry out intricate sequences of manipulations when building structures, or do repair work. Great apes, like chimpanzees and orangutans, are also capable of bringing tools with them, even when they cannot directly observe the target site where the tool will be of use (Osvath & Osvath, 2008). Similarly, Kabadayi and Osvath (2017) showed that ravens can apparently recognize the future use of a tool for a time span of 17 hours, and will also barter to get hold of such a tool.

Buzsáki and Tingley (2018) proposes that the brain cannot represent time directly, but must do so by means of sequences akin to movement. If true, this hints at a way by which machinery for spatial processing discussed in the previous chapter can integrate with skill–based manipulation abilities to support sophisticated behavioural sequences, involving both spatial and temporal ”places”, as well as place–specific tasks.

Tasks tend to associate functional behaviour with an outcome. By disassociating outcomes from functional context, and vice versa, a powerful cognitive ability comes into being. Next we will look at the manipulation–state metaphor.

7.7 Machinery for manipulation provide abstractions of function and abstractions of outcomes: the manipulation–state metaphor

In chapter two, I looked at sensory generalizations as the down–regulation of detail information, that affords recognition based on similarity, rather than equality. The generalizations that will discuss here are more powerful though, since they in effect afford *transfer learning*.

In the context of manipulation, outcome abstractions correspond to a focus of attention on the end product. This corresponds well to earlier investigations into how goals may be induced by interoceptive prediction errors, and the awake cognitive system being tasked with coming up with a way of getting to a site where the prediction error can be reduced. Similarly to approaching such a site requires knowing the way there, the production of a manipulation outcome requires the manipulator to know the production process. Such processes may be learned by observation, by being taught, or by trial and error. Hence a crow can learn how to crack nuts by placing them in the way of cars, or it may discover that dropping the nut on the road may also do the trick. Discovering the underlying kinetic principle of a process, will in turn afford synthesis of new processes. This can yield increased allostatic flexibility.

Outcome abstractions are complemented by *functional abstractions*, the notion that the same manipulation procedure can be reused to produce a range of different outcomes, and in different problem contexts. One way of testing this ability in animals, is by the use of trap–type tasks. In these tasks, the animal must push or pull food in the correct direction, otherwise the food will fall in a trap and become inaccessible (Martin-Ordas & Call, 2009; Martin-Ordas, Call, & Colmenares, 2008). The ability to use skills for completing one such task in a different one is usually referred to as transfer learning. There are many indications that transfer learning is effortful, also for humans (Jonsson, Kulaksiz, & Lithner, 2016; Salomon, 1992).

In the comparative cognition literature, the term physical cognition is used to refer to the prediction of movement due to physical law, as well as the prediction of effects of action in this context. In Tjøstheim and Stephens (2021), we argue that manipulation of working memory content, such as rotation or translation, can be used for vicarious trial and error, similar to its use in navigation. Also similarly, trying out manipulation in working memory can save energy on trial and error and in some cases allow the animal to directly predict what behaviour is required (Heinrich & Bugnyar, 2005; Kabadayi & Osvath, 2017). However, manipulation skills open up larger vistas, in the sense that it can support the fashioning of specialized tools (J. Lewis, 2006). Hence it is the foundation for the development of technology. As human history shows, this ability again affords large scale manipulation of the environment, and the construction of niches (Odling-Smee, Laland, & Feldman, 2003).

7.8 Summary

In this chapter, I have added to the concept of places the affordance of manipulation and change. Some places stay the same, while other places change frequently. Animals that live in places that are variable experience pressure to generalize, and to be less dependent on particular foods and constant living conditions. Hence both perception and behaviour must become more flexible to cope with variability. Many carnivorous organisms are both predators and prey. This means that their ecological niches tend to be more complex than that of herbivore organisms. This is so, since predators have to predict the behaviour of their prey, but prey organisms can actively behave to foil attempts to be captured and eaten. Usually predictive abilities have an energy cost, but this can be compensated for by the higher energy density of meat compared to plants. Accurate prediction is dependent on precise control of neuronal behaviour. This is aided by myelination, as well as contribution from cerebellum, which can learn to predict and coordinate complex phase dynamics in the rest of the brain. The ability to independently and precisely control fore–limbs opens up abilities to manipulate and actively change the environment, as well as to pluck and carry food. Foods that are particularly tricky to get at require the development and learning of skills. Under some circumstances, like getting ants out of a ant–mound, or nuts out of shells, skills can include the use of tools. Like food, tools may be carried from one place to where they are to be used. In cases like this, the flexibility of the navigation–place machinery may be appropriated to plan, allowing useful tools to be brought along on foraging trips, even if they are not immediately useful. As manipulation skills become well developed and automatized by the cerebellum, they can be used like vicarious trial and error, to mentally manipulate objects in the environment.

Chapter 8

Towards an artificial cognitive system

How might the concept of a cognitive system be concretized into artificial machinery? Are the processes of allostasis necessary in machines, perhaps as a means to secure electrical energy? Further, can cognition as realized in animals be abstracted from their allostatic context and meaningfully transferred to artifacts? Earlier in this thesis I have discussed various sources of motivation that are related to energy saving, if not directly to energy gain. Curiosity and novelty can induce motivation to explore and to gather knowledge.

Before attempting to answer the questions posed above, I will fill in some detail about the simulations conducted as part of this thesis. In essence these simulations have been carried out at either the system level, or at the level of individual neural units. In either case, they can be thought of as computer programs that are run many times, but with some change in parameters and input values. This process may seem obscure to someone not familiar with computer programming, and even those that are may wonder at the process. It may help to look at it from a few different perspectives, and using some analogies. Creating the model itself is much like creating and writing down a *recipe* for a meal, compared to actually cooking the meal. Cooking the meal would in this case be like *running* a simulation. In the same way that you can cook many meals with the same recipe, but each time you use different potatoes and different carrots, different simulations can be run with the same model, but using different input values such as for example images. Changing the parameters of the model is similar to changing the *amounts* of ingredients in the recipe without changing which *kind* of ingredients, or how you add them together: you can make a vegetable stew with one teaspoon of salt, or you can use two. The recipe is quite similar in either case, but the meal could taste quite differently.

Writing recipes can be done in several ways. You can use a pencil and a notebook, and you can do it with a typewriter and sheets of paper. Also, you can borrow parts of other recipes and add them to your own. Imagine that you want to make a vegetable pie, but you only want to specify which vegetables go into the pie. You can borrow the recipe for the base of the pie from a book, and copy it into your own recipe. In the same way when creating computer models, it is possible to write them using different *frameworks*. We have used two such frameworks, called Ikaros (Balkenius, Johansson, & Tjøstheim, 2020), and Processing (Reas & Fry, 2007). They both have strengths and weaknesses, and may make different things *convenient*. Ikaros, for example has lots of ready made recipe parts that can easily be stitched together, almost like Lego. The Processing framework makes it easy to visualize and draw things. Either way, when a model has been programmed, it is run in a a sequence of *steps*. Although a recipe may have steps as well, the computer simulation version of steps is more like frames of a movie, and the model describes what is necessary to create the frames.

Three of the simulation models in this thesis (Tjøstheim & Balkenius, 2019), (Balkenius et al., 2018), and (Balkenius, Tjøstheim, et al., 2020) were done at the *systems level*. This means that we did not simulate the details of individual neurons, or even populations of neurons. Instead, we focused on the behaviour on some assumed network in the brain, and calculate more abstract activity levels. This typically boils down to calculating a list of numbers, and each number is usually between zero and one. These numbers may be interpreted loosely as average and normalized activity of many neurons, but is focused on what kind of function the neurons do, rather than exactly how they do it. For example, in Balkenius et al. (2018) one function is to provide a kind of association memory between different sensory inputs, while another function is a memory of sequences.

In Tjøstheim et al. (2021) we simulated at the level of spiking neurons, using an adaptation of the algorithm presented in Izhikevich (2003). We also made the simulated neurons control a little agent that could move around in world, similar to that of a video–game. The agent sent images of what it saw into the simulated neurons, thus completing a kind of motor–sensory loop. Compared to the system level simulations, this one is more like a real biological system, even if it is still very simplified.

I have argued in this thesis that cognition can be understood as a tool for allostasis, but the simulations described above have no physiological or allostatic aspects. What is more, cognition has been studied for many years with little impediment from ignoring allostatic processes. Next I will look closer at this apparent paradox.

8.1 Abstracting cognition from the allostatic context

In biological systems, allostatic processes fundamentally dictates *goals* to the organism, that is places and things to approach or avoid. Nutrients and mates are examples of things to approach, while predators are examples of things to avoid. The role of the cognitive subsystem is then to guide the organism to these goals, and negotiate any obstacles that may stand in the way. What is more, allostatic processes naturally signal to the cognitive system when the goal is reached, both in terms of when e.g. a source of nutrients is found, but also in terms of consummation. Satiation is a physiological signal that consummation should terminate. To borrow a term from computer science, allostasis solves the *halting problem* for the organism. The halting problem was formulated as the problem of predicting whether an algorithm would stop running or go on forever, given only a description of the algorithm and its input data (Church, 1936). Physiological limits typically has both a lower and an upper value, as we saw in the case of calcium homeostasis for neurons. If values move beyond these points, it typically reduces wellbeing for organisms, and triggers compensatory behaviours. In any case, if the values move too far from the physiological range, the organism will eventually disintegrate, and all processes will naturally halt. The aging process is another way that organisms are naturally ensured to eventually stop behaviours.

In technological systems, and in absence of allostasis, stopping conditions must explicitly be programmed in. Artificial systems also lack natural goals, though they can be given goals or problems to solve by us. However, this situation gives rise to problems well known to artificial intelligence (AI)researchers of how to specify the *utility function* of an intelligent agent (Bostrom, 2014). The ”utility function” for biological systems is precisely allostasis; reward comes from moving the system closer to homeostasis, for example by eating energy rich food when hungry, or drinking water when thirsty. In an AI system this problem is compounded because we don’t want the such systems to hurt us or somehow work against us. That is, we want the goals of the AI system to be *aligned* with our own wellbeing. This means also that even if we were to give the system abstract and high level goals such as ”end world hunger”, there is a lot of implicit constraints in that goal that we want the system to understand (such as e.g. not killing everyone) and take into account as it works towards a solution.

There is a difference between science fiction–like autonomous and intelligent robots, and using robots as a research tool to investigate cognition and cognitive processes. In constrained laboratory settings, it is readily possible to divorce cognition from allostasis. But it might be instructional to use the regulatory aspects of allostasis even in artificial systems, to drive behaviour. In the next section I look at some ways this might be done.

8.2 How can allostasis be interpreted in a context of artifacts?

A computer plugged in to an electrical socket has a constant source of energy. What is more, computers have memories that can keep their information, even if the electricity runs out. Even a laptop that relies on a battery won't be destroyed if that battery runs out. In these respects our computers are fundamentally different from biological organisms, who quickly die and disintegrate if their energy supply runs out, and whose memories and experience too can be destroyed if the flow of energy and oxygen should be disrupted.

But even if our everyday computers do not by themselves require an allostatic system, nor depend on allostatic behaviour to keep them going, we can still program them to have propensities similar to biological organisms. Already, lawn mower robots and vacuuming robots are routinely programmed to find their way to a charging station before their batteries run out (Silverman, Nies, Jung, & Sukhatme, 2002). This does not imply that those robots have any kind of feeling that is similar to organisms. Nonetheless, if a robot were constructed with an ability to plug itself into any kind of electrical outlet, as well as a visual system to recognize such outlets, along with a spatial system to learn where outlets are found, this would be steps along the way of creating a system that has interesting analogies to biological systems.

Pathak, Agrawal, Efros, and Darrell (2017) trained a system of neural networks to play the games *Super Mario Bros* and *VizDoom* using three broad settings: first, sparse extrinsic reward on reaching a goal; second, exploration with no extrinsic reward; and third, generalization to novel scenarios. In these experiments, the authors operationalized intrinsic reward from success in being able to predict the consequences of its own actions in the game. According to them, they were able to avoid some difficulties reported by other experiments by filtering away data from the input that were not due to the agent itself. They found that this agent could learn quicker in the first condition compared to an agent that did not use intrinsic reward, and that in the second condition without any reward from the game, the agent could still make progress. In the third condition, the authors showed that an agent with experience from an earlier level could outperform one that did not have this experience.

Using a robot finger with touch sensors, and operationalizing curiosity–driven learning, as minimizing computational resources spent on representing sensory signals, Pape et al. (2012) show that this system can develop a set of skills that yield particular tactile signals. They report that the system starts out with random exploratory behaviour, but since the system is set up with the curiosity–like algorithm described above, it will gradually end up with this small set of skills. The researchers then show that the system can exploit these learned skills to do texture classification.

8.3 What might thinking entail in a machine?

To get an inkling about what it means to be at all *thinking*, Holyoak and Morrison (2012) give the following six examples. One is the expression of *beliefs*, like ”I think there is someone at the door”. Another is *problem solving*, like ”I hope you can think of a solution”, which the authors also associate with finding a path to a goal. A third is *foresight*, and imagining or predicting the future, like ”I think that tree will fall next summer”. Fourth is *judgement* and assignment of value, like ”did you think the movie was any good?”. Fifth is a variant of judgement involving *morals*, like ”I think lying is wrong”. Sixth is something like *daydreaming* or *reflection*, like ”He spent the afternoon absorbed in thought”.

The question now becomes whether a machine would be able to plausible produce these kinds of mental behaviours. From the preceding chapters, I would argue that most should be in principle possible for a machine, even if the resolution might be low. Having said that, some philosophical charity is required to accept the operationalization of the various psychological terms, like belief, foresight, and daydreaming, in machine terms. Also, communicating about thinking is a separate issue which is beyond the scope of this thesis. I will attempt to justify these claims in this section, by looking closer at each of the six examples in turn.

A *belief* in machine terms may be operationalized as a probability distribution of salient events. To unpack this, we can recall some concepts from chapters five and six. Given that we have implemented a perceptual network that supports selection of percepts, depending on contrast and importance, that is depending on salience, and that these percepts are channeled into a sequence memory, which has the property of grouping together sequences according to their statistical similarity, these statistical, or probabilistic sequences should be analogous to beliefs about causal sequences. For example, imagine a robot repeatedly perceiving the following sequence of events; first the sound of knocking on a door, then the door opening, then a person behind the door; by association, this sequence of events will be likely to play back in the memory network of the robot. To further make up a belief that it was the person doing the knocking, the robot also has to repeatedly observe a person actually doing knocking, such that the knocking sound is associated with the visual perception of the person using her arm to knock. The point is that the playing back of these associative sequence are analogous to, and an approximation of thinking in the sense of ”I think there is someone at the door”.

Earlier, I argued that abstract problem solving may be operationalized by means of the manipulation–state metaphor, where typically hands are used to change the environment such that approach behaviour is facilitated. In robot terms, problem solving like this might be concretized as moving something like a cardboard box out of the way, such that the robot can move from one place to another. Although this may seem a trivial problem, the interesting part is if the robot has experienced similar enough situations, where it uses its arm to push or move things, that it recognizes the possibility of engaging in actions, like pushing or lifting, to move the box. The cardboard box example is perhaps simple enough that reinforcement learning could be used. This implies the association with some percepts with particular actions in the manner described above for artificial curiosity. The essential part of this kind of problem–solving thinking is the simulation of event sequences that involves both actions and the effect of those actions, as well as a state of the world that serves as a regulatory set–point that is to be worked towards.

Foresight, in the context of machine thinking, is similar to problem solving, but is in some ways simpler. Depending on the robot's experience aggregated in its sequence memory, foresight would entail simulating possible event sequences given the current observed state of the world. In the example ”I think that tree will fall next summer” this would imply that the robot has similar long term experience as a human, spanning years. This could perhaps be simplified by using the cardboard box example, if the box is placed in the middle of a corridor, as ”I think that box will be moved”, in the sense that repeated observations of people moving the box out of the way would trigger the playing back of such an event sequence on observing the box.

Leaving aside the judgement of good and bad, as well as right and wrong, we are left with daydreaming and reflection, in the sense of ”He spent the afternoon absorbed in thought”. These two situations may be differentiated by reflection being constrained to some specific subject, while daydreaming is not. Daydreaming could then be operationalized as letting a sequence memory and associative memory freely activate across any learned content. This is in essence what Balkenius et al. (2018) is about. Restricting memory to a particular situation would entail entraining the memory by e.g. forcefully driving it by means of executive top down activations (Gu et al., 2015; Smallwood, Brown, Baird, & Schooler, 2012) and downregulating noisiness in the network to keep it from moving along weak associations, and away from the desired situation.

8.4 Simulation and its limitations

As a scientific tool in cognitive science, we would like computer simulations of both overt behaviour, and of cognitive processes to inform us about corresponding behaviour and cognitive processes, in animals and humans. There are various ways this may be the case, carrying different degrees of scientific value. One type of simulation makes no attempt at predicting biological systems, but only shows that some function is *possible* to implement on a computers. For example, that it is possible for a robot to recognize a person and signal that it has done so. Many classical AI systems relied on computer science, and did not attempt to mimic biology in any way. They contributed by making concrete what kind of computations could be involved in realizing a function like recognition. A second type of simulation aims for *qualitative* similarity, for example that the plot of simulation values looks similar to the plot of recorded values from an animal. In Tjøstheim and Balkenius (2019) we present a plot of how visual acuity develops in kittens, and show that the plot of how similar an input image is to a reconstructed image in our systems qualitatively resembles the kitten plot (i.e. it converges over time to some stable value). A third type of simulation uses experimental data to *fit* the parameters of the simulation, such that the simulation reproduces the experimental data as closely as possible. This method can also be used to select between different models, to see which of them produces the best fit. For example, when simulating single spiking neurons, data recorded from a living neuron, like electrical voltage and current over time, may be used to adjust the parameters of a simulation of action potentials such that they are as similar to the real thing as possible. This simulation can then be applied to reasonably predict what would happen in biological neurons under different conditions.

Using experimental data to fit simulation parameters can yield valuable scientific knowledge. But a limitation of this approach is the danger of *overfitting*; this means that the parameter values fit the testing data very well, but that they fail to generalize, and predict data from different, unrelated experiments. This is particularly a risk, when models have many parameters, and the risk grows with each added parameter (Dietterich, 1995). In the example about spiking neurons, it may thus be the case that the simulation can reproduce the rhythm of e.g. pyramidal cells in the prefrontal cortex very well, but completely fails to reproduce the rhythm of the same kind of cells in the parietal cortex, even if the input is the same.

The legitimacy of science rests among other things on whether it is possible by arbitrary researchers to accurately reproduce experiments. For computational simulations, being able to reproduce or re–run experiments may bring many challenges, which can limit the value of simulations as scientific tools. Concretely, computer simulations are software, and may not directly run on computers that are different from the ones the original researchers used. Some times the simulation software must be compiled from the source code, which can bring its own difficulties, like finding and installing required software libraries with the correct version. It may also be a challenge to understand what the simulation actually does, or how it works, just by looking at the source code.

Physical experiments are the gold standard in science, and typically have the highest scientific value. Simulation experiments compare to physical experiments in that they can yield numerical data which can again be subject to statistical processing and interpretation. Similar to physical experiments they may also be time consumptive to set up and run, hard to replicate, and prone to failure. Simulations contrast with physical experiment in that they do not directly ”ask nature questions”. That is, simulations have a different degree of assumptions and bias built in to them, since they are from the bottom up constructed by researchers. On the other hand, even experiments on animals may impose on those animals unnatural conditions and place them in situations that are very different from what they have evolved to cope with (Cole & McDermott, 1997). Simulations, once programmed, afford the possibility of parallel scale which is harder to do with physical experiments on animals and humans. That is, the same experiment may be run in parallel with different parameters on a computer with comparatively little cost, thus yielding a broader look at the landscape of possibilities. Simulations also afford control over parameters in a more precise and stringent way than can physical experiments. For example, in the agent simulations we conducted in Tjøstheim et al. (2021), we varied the number of simulated neurons available to the agent. Adjustments like these are usually not possible to do with real animals.

Simulations compare to philosophical theorizing in that it is often based on idealization and simplification, abstracting away the noise of nature. In a way, computer simulations are constructing philosophical arguments using the computer as an active tool. Philosophical arguments rely on logic, and logic is also a necessary part of computer programs, including simulations. Similarly, philosophical deduction, starting from some limited set of axioms, can expose consequences of those axioms; in the same way axiomatic assumptions are used in simulations, and the running of the simulations will expose the consequences of those assumptions. Computer simulations contrast with philosophy in that they afford embodiment both in virtual agents in virtual worlds, but also in physical robots acting in the real world. Numerical simulations involving probabilities and dynamical systems allow a much higher complexity than do philosophical arguments relying only on logic, and the limits of human working memory. In this way, computer simulations allows bridging philosophy and physical reality.

8.5 Summary

In this chapter, I have discussed how cognitive processes may be implemented in artificial systems that are not reliant on allostasis. However, by removing allostasis, cognition is also deprived of natural goals, as well as a substrate for value and judgement. Artificially providing goals that can drive complex behaviour is not trivial, and neither is replacing physiological value. However when considering foraging for information in the form of novelty and knowledge, artificial systems could likely be made similar enough to biological ones that complex behaviour could emerge, including mental behaviour. Thinking is a broad term for many kinds of mental behaviours, that can include maintaining beliefs about causal relations, problem solving, foresight, and reflection. These behaviours have in common that they rely in part on association memory and sequence memory, both of which can be implemented in machines. Understanding thinking this way means that it may not be impossible to construct thinking machines. Creating such machines will entail running many kinds of computer simulations, and these kinds of simulations are also scientific tools within cognitive science. Compared to physical experiments, computer simulations have less direct validity, but are flexible and allows for precise control of parameters. Computer simulations have flexibility and freedom in common with philosophical thought experiments, but can be made more strict and more complex than thought experiments. By letting simulations control virtual agents or physical robots, simulations may be thought of as bridges between philosophical speculation and physical experiments.

Chapter 9

Contours of cognition

In this final chapter I attempt to re–tread and summarize the path I have taken, from thermodynamics and single cells, to foraging, navigation, and obstacle negotiation. Zooming out, what are the contours of cognition, and what is the interplay between environment, and the shapes and forms that cognition might take?

9.1 Cognition as a tool for allostasis

In chapter three and four I talked about regulation and allostasis, and that allostasis is the name for all processes in the body that work to keep the organism alive. Inside the body, allostatic processes are often referred to as *physiological*. They regulate the amount of oxygen and sugar in the blood, and carry away waste products including carbondioxide. But many organisms need to move in their environment to stock up on energy, water, and building materials. In these cases, behaviour is allostatic in the sense that it is necessary to be able to find the needed resources. But to be able to find the resources, the organism also needs to process information from sensors, and engage its muscles to create movement towards the place where the resources are. The processes that are involved in transforming sensory signals to muscle signals are cognitive processes, and just like the movement behaviour, and behaviour of the internal organs, they are allostatic too: they contribute to keeping the organism alive.

The specific nature of neuronal cells and the way they are connected together produce an interesting kink in this pattern though. As we saw in chapter four, neurons are somewhat delicate in that they are dependent on admitting calcium at a relatively narrow rate: not too little, and not too much. Otherwise they stop working and fall apart. The brain uses patterns of inhibition to represent what it expects will happen, so that the sensory signals that match the expectations will be cancelled out. But if something differs from the expectation, it will be propagated onward. The kink is that if the predictions are too good, less than the optimal amount of calcium will be let into neurons. At the subjective level, this can be felt as boredom, and can motivate people to seek out novel sensory impressions. In practice, this can create an urge in people to get out of the static office and into ever–changing nature, but can also motivate them to travel to places they haven't been. Once again both behaviour and cognition can be seen as allostatic, since both of them are necessary to find and move to place that can give new sensory impressions.

Larger neural populations can typically cope with predicting more complex sensory impressions, but on the flip side such populations require more variation and complexity to work properly. In the next section, I will look at how cognition may be understood on a scale from simple to complex.

9.2 Cognition across a spectrum of complexity

In this thesis, I have argued that recognition and choice are fundamental functions of cognition, but these are both quite broad categories. Hence in this paragraph I intend to consider systems that very likely are not cognitive, and why it is that they cannot support recognition and choice. Again I make use of the terms *orthograde* and *contragrade*, which I remind the reader mean going with a natural flow, or going against it by using energy. I will begin with extremely simple and obvious systems, and gradually try to increase complexity to get closer to a minimal cognitive system. First, consider systems that lack internal structure and boundaries, like gases at equilibrium. For such gases there can be no allostasis, and there is nothing that either do recognition or choice. Next consider a crystal structure like a diamond. A diamond can display a clear boundary and has an orderly structure. But this structure is too rigid to display behaviours, although it can be affected by thermodynamical perturbations, and can be heated or cooled. A body of water is next. Water can have rich internal dynamics, but unlike a diamond, it needs some external structure to contain it or it will spread out, and may turn into gas. However, a body of water in the presence of a gradient will display some self–organizing tendencies as e.g. described in Deacon (2011); Bernard cells can form in thermic gradients, like water on a stove; vortices can form in gravity gradients like water running in a river towards the sea. These dynamic phenomena will also tend to preserve their dynamic structure as long as the gradient persists. But Bernard cells and vortices are still orthograde; there is nothing in them that allows their pattern to oppose the gradients they are exposed to. Neither is there any way for either a Bernard cell or a vortex to recognize anything or to exercise options.

Let us recall the slime mould from earlier chapters. The particular way in which the slime mould explores its environment, by spreading out and compressing itself around nutrition sources, means that it does not necessarily orient itself as do most other organisms. The slime mould grows in every direction, and does not have a particular body shape. This is particularly noticeable when we observe how the slime mould behaves if it cannot find enough nutrients. By forming a fruiting body filled with spores that is released into the air, it uses atmospheric pressure gradients to spread itself out over an extended area, in a similar way to how it spreads itself out into its local environment. In this way, it is perhaps more similar to a plant, than to animals. We can keep this distinction in mind when we now turn to single celled and multicelled organisms.

Bacteria like *E. Coli* have sensors that can detect chemical gradients, and that are coupled to effectors that can spend energy to move the bacterium along the gradient toward nutrients. This system has an enclosure or cell wall, and can apparently recognize stimuli in the form of chemicals. Although it can move it has limited choice (Greer-Phillips, Alexandre, Taylor, & Zhulin, 2003). In particular, *E. Coli* cannot choose to move away from a nutrient gradient. We will consider two more organisms: first, the *C. Elegans* a small nematode worm that lives in soil and eats bacteria; second, the sea squirt which has a central nervous system in its larval stage, but reduces it later.

First, *C. Elegans* is interesting because it is multicelled and has a simple nervous system consisting in less than 400 neurons and around 5000 synapses (White, Southgate, Thomson, & Brenner, 1986). This organism has a range of sensors, including ones for chemical concentrations, and for touch. It has also a muscular system that lets it both approach and avoid stimuli in its environment. In addition to recognition, *C. Elegans* appears to be able to make simple choices, both for what to eat (Shtonda & Avery, 2006), and how to move (Shinkai et al., 2011). Although choice behavior is difficult to separate from reflexes (Boyle, Berri, Tassieri, Hope, & Cohen, 2011), C. Elegans appears at least to have more options than do *E. Coli*.

Second, the sea squirt is salient due to the changes that happens to its central nervous system over its lifetime. The sea squirt starts out as a mobile tadpole with sensors for light and gravity (Ryan, Lu, & Meinertzhagen, 2016). In this life stage, the sea squirt can swim and orient itself, and seek out a place where it can settle. Once it has settled, it stays in the same place for the rest of its life, and most of the nervous system is transformed into a more developed digestive system including structures for capturing nutrients from flowing water by filtering. The tadpole central nervous system consists of 177 neurons (Ryan et al., 2016), and is thus a little more than half the size of the C. Elegans nervous system, but has around six thousand synapses compared to the C. Elegans five thousand. Although little appears to be known about choice or decision processes in the sea squirt tadpole, it is likely that it can recognize places to settle.

The contour we can spy when comparing the non–living systems with the living systems, and the living systems with each other, is the notion of *space*, and how organisms cope with spatial aspects, like gradients and places. The slime mould blankets its limited local environment by progressively spreading out in it, or by releasing spores. The oscillators that it is made up of, reacts to nutrients by cycling faster, and to noxious substances by slowing down. The sum and interaction of these processes support allostasis in the slime mould. Bacteria like *E. Coli* are focal in shape, can orient themselves, and move along nutrient gradients. If no nutrient gradient is recognized, *E. Coli* will randomly move about in its environment, until recognition happens by molecules slotting into its chemical receptors. The *C. Elegans* worm hunts bacteria at least in part based on olfaction (Worthy, Rojas, Taylor, & Glater, 2018). That is, like *E. Coli*, they can orient themselves, move along chemical gradients, find, and ingest nutrients. *C. Elegans* appear also to be able to develop preferences, and have been observed to avoid some bacteria based on experience (Shtonda & Avery, 2006). Finally, and perhaps most saliently, the sea squirt demonstrates when, and how, cognitive processes are necessary. In its tadpole stage, its nervous system performs the occasion–limited function of seeking out, and navigating to a place to stay. When it is active, the neuronal processes of its nervous system supports orientation in relation to gravity, and in relation to light, and can recognize suitable settling places (Ryan et al., 2016). When settling is complete, and coping with space is no longer required, the nervous system is dismantled.

The sea squirt's time as a cognitive being is limited, and presumably its need for memory is limited too. But *C. Elegans* apparently learns from experience and uses it to avoid some food, and approach other (Shtonda & Avery, 2006; Worthy et al., 2018). In the chapter about foraging I talked about space, but also about how foraging can be supported by learning. That is, how an organism can save resources by recognizing sensory sequences that repeat. This ability allows prediction in particular. We can recall from the chapter about regulation that, the organism itself is also a source of repetitive patterns that must be predicted for physiological processes to do their allostatic work. While foraging though, prediction of visual impressions in particular, are helpful for an animal that is making its way towards a place with food or water. Similarly, accurately predicting processes, involving gravity, and such mechanical processes as bending, and properties like flexibility, makes it possible for an animal to use tools, and predict how to use the tool to get food. Prediction of physical processes like this, combined with prediction of how own behaviour can modify the environment is the machinery that mediates manipulation, as we saw in chapter 7.

From this we can recognize a second contour: coping with repetition and change by learning, which again affords *prediction*, and predictive regulation.

9.3 Reflection and thinking can integrate and associate knowledge from different sources to make better choices in future

When an animal is in the midst of a foraging situation, goal directed, focused prediction and behaviour, means that immediately relevant associations are most likely to be activated (Shettleworth, 1989). However, when an animal has returned to its nest, when immediate needs have been satisfied, and it is safe, the default mode network takes precedence over the seeking and executive networks (Greicius, Krasnow, Reiss, & Menon, 2002). In a relaxed state, and when no longer focused on approaching a goal, attention is typically turned inwards, and attentional focus will tend move around more freely since the winner–takes–all mechanism described in chapter 5 is also relaxed. In this state, populations with weaker activation are more likely to win out, and become the attentional focus.

The default network includes areas of the brain that processes both spatial and temporal context, places and sequences, as well as action choice and associations of various kinds (Greicius et al., 2002). When this network is left to spontaneously roam, the result is a kind of daydreaming (Kucyi & Davis, 2014). Associations that are followed can include walking around known paths, since each place along a path are associated as the animals traverses it. But since the grid system is activated at the same time, it is possible for an animal to become aware of possible shortcuts, or alternative paths between places, that it has not physically experienced. Similarly, different sequences of manipulation actions that lead to the same outcome, or outcomes that have something in common, can form the basis of generalization. The noisy aspect of the default mode, and its propensity to admit associations that are specifically weak, are key in the sense that it allows an organism to become aware of non–obvious commonalities.

Another aspect of default mode activation is the re–experience of particular trains of events, in contrast to the free association described above. We can call this *reflection*. The reflective mode allows imagination of possible alternatives of behaviour, and what might be the outcomes of those alternatives. For a foraging animal, this kind of reflection may only concern journeying to and from foraging sites, or possibilities for exploration (Poe, Nitz, McNaughton, & Barnes, 2000; Redish & Touretzky, 1998). For social animals, there are indications that this behavioural–outcome simulation also includes social interaction (Li, Mai, & Liu, 2014; Mars et al., 2012).

Even if the default network can come up with alternative scenarios and hypotheses about possible paths to try, or behaviours that might give better results, how reliable are these default processes (Kornblith, 2012)? While it is clear that direct experience is more reliable than scenarios that occur when mind–wandering, there are some specific situations where an animal may benefit from having the ability to mind–wander. In Stephens and Tjøstheim (2020), we argue that in novel situations, a strict reliance on direct experience will in fact be a liability, since in nature change is constant, and although patterns are profuse, nothing ever repeats exactly. If the novel situations bears some resemblance to situations created during reflection or mind–wandering, the animal has some behavioural option instead of merely freezing up. In this way, reflection can contribute to reliability of mental models, by means of corroboration and association of independently acquired experiential traces.

9.3.1 Foraging and exploration in conceptual spaces

So far this thesis has been chiefly concerned with cognition in the context of non–human animals. In this section, I will in contrast only talk about human conceptual cognition. Human advanced metaphors can be used to compose concepts on the fly to aid choice of behaviour in approach to goals, and also to choose goals. These goals can include states of the world, but also allostatic states.

The notion of *conceptual spaces* is due to Gärdenfors (2000) and proposes that human categories are spatially organized by similarity. Here I will use this term in a somewhat broad interpretation, where concepts are simulation models in the sense of Barsalou (2009), but are not necessarily distributed in conformance to Cartesian axes. However, concepts may still be seen to be spatially organized in a place–like fashion, such that there may be mental paths between places, and places contain classes of concepts, each of which can be *manipulated*.

In this way, foraging in conceptual spaces can mean using and synthesizing concepts to concretize a goal in the form of a desired outcome; to find a path to that desired outcome; and to move along that path. Oftentimes the path is uncertain, and requires extensive exploration to learn conceptual models, such that they can be used when doing mental path wandering and manipulation, in the form of simulation. Finding particularly energy saving metaphors and concepts map to places of tasty and satisfying food; finding a path between groups of concepts is motivating and exhilarating, as it allows us to move between conceptual places that we before assumed were not connected. Our rich mental landscape of connected places, where we can do and make different things, affords us rich abilities to predict the natural world, to predict our social group and society, and crucially to shape it such that it can save us energy.

In the following I will propose some hypothetical conceptual metaphors, that make use of various intrinsic processing machinery that human brains have evolved. I start with metaphors of space, since above I identified space as a contour of cognition, and indeed most of this thesis concerns space and navigation. Chapter six in particular was about the navigation–place metaphor, and I will provide some more examples here. First, a journey involves both navigation and places. A journey often has some particular place as a goal, but many journey start and end at the same place, our home. When we come back from a journey, we tell people about it; the journey can be made into a *story*. Particularly interesting parts of a story are those where tension and uncertainty arises, often when an obstacle occurs. The audience wants to know, which option did you choose? Which path did you take? How did you get away from the bear? If the journey had no tension, it can still have aspects of novelty. This is pleasing too; getting to know about how people behave in distant places, what the houses look like there, what they eat, how the landscape is; all of it generates interest, since subconsciously it has potential value for the listeners: it would help them predict and navigate around that place if they were to go there themselves.

The interesting point is that, even stories which are not about journeys, still appear to use the machinery of spatial cognition. Milivojevic, Varadinov, Grabovetsky, Collin, and Doeller (2016) conducted an experiment where participants watched parts of a movie called *Sliding doors* while placed in an fMRI scanner. This movie consists of going back and forth between two narratives, depicting alternative outcomes as the protagonist either misses, or catches, a subway train. The data recorded in the experiment indicated that neural activity in the hippocampus, which is responsible for processing places and sequences, was different when the movie showed particular people and locations. The results indicated that the hippocampus also differentiated between the two story lines. This study thus gives support for metaphors like ”life is a journey”, and ”relationships are journeys” being mediated by areas in the brain that also process concrete places and navigation.

Another aspect of space processing that we saw in chapter six, is that of the grid system, which allows us to use map–like representations, and to make inferences about how places are spatially related. As it turns out, this system is versatile, and can be used to process even abstract relationships. Garvert, Dolan, and Behrens (2017) tested this notion by showing human subjects a series of random images, including a motorcycle, an umbrella, and a shoe. The subjects were told that the images were randomly picked, but in reality, the images were placed on a grid such that for example, the shoe was connected to the umbrella and a picture of a rabbit. Instead of the sequence of images being random, they were always shown in a way that corresponded to walking along the connections of the grid. In effect, the images acted as ”places” or locations, that were connected via other places. The researchers then showed the images again to the subjects the following day, but now the subjects were placed in an fMRI scanner. The activity of the subjects' brains, that is their hippocampus and their entorhinal cortex, displayed activity that reflected the spatial–like organization of the original grid, even if the subjects never saw anything but the sequence of images, which they also thought was random.

As we saw in chapter seven, sometimes it is not enough to travel to a place to get what we need; sometimes it is necessary to change the place itself. For example, we might a found a nice place with a fine view over a lake, but we still need to build a house there to have shelter and somewhere to live. On a large scale, we might want to use the energy of a waterfall to drive turbines and create electricity. On a small scale, we need to solve problems of many kinds, from writing reports, calculating income taxes, cook a dinner, or tidy up the living room. These are all examples of manipulation, in the basic sense that we typically use hands to change something into something else: we change an empty sheet of paper into one filled with sentences and paragraphs; we change the income tax form from having a lot of empty fields, to one where numbers are written that reflect our income and expenses; we change five potatoes, some carrots, and onions into vegetable stew, and we change the living room from a mess of toys and magazines and used plates, to a room where every thing is in its place. Even more abstractly, a mathematician might change an equation into a different form as part of a proof, a computer programmer might change a routine of computer code to fix an error or to transform a list of numbers.

Manipulation activities that change the state of the world is primarily associated with areas of the parietal cortex. Alivisatos and Petrides (1997) placed subjects in a positron emission scanner and took pictures of their brain activity while they were tasked with telling the difference between numbers and letters and their mirror image. Subjects were also told to discriminate between rotated numbers and letters. These tasks required subjects to change the stimuli before comparing them, in effect making the subjects either do a mirroring transformation, or a mental rotation before answering if stimuli was the same or different. The brain images of these subjects were then compared by the researchers with control subjects that did comparison, but did not have to do transformations. The researchers found that areas of the parietal cortex were activated in the experimental group, while less so in the control group.

Rotations and mirroring transformations can be done by the hands on physical objects. But changing sounds and melodies are less obviously a candidate for handed manipulation. Zatorre, Halpern, and Bouffard (2010) tasked musicians placed in a fMRI scanner with comparing notes of a familiar tune with a correctly reversed version, and one that had errors. Like the mental rotation experiment above, this required the subjects to mentally transform the note sequence by reversing it, before making the comparison. The results from this experiment showed that an area of the parietal cortex called the intraparietal sulcus was particularly activated, along with more frontal areas that tend to be involved with memory. The researchers propose that parietal areas could be involved with quite general transformations of sensory input, not only visual input.

As the above experiments indicate, there appears to be support for the notion of spatial metaphors. On the one hand, it seems that we can mentally move in conceptual space, as long as there is some kind of spatial–like structure that the hippocampus and the grid cell areas can latch on to. This enables us to make inferences and pose hypotheses about domains that we cannot directly observe, or that are abstract. Physics and mathematics are examples of such domains. On the other hand, when we find ourselves at a particular "place" in conceptual space, it appears that we can use areas of the brain that likely evolved to support the control of the hands, to mentally change information from the senses. Together these abilities make for powerful abilities to both imagine and change our environments to save energy and improve quality of life for our species.

9.4 Summary

In this chapter, I reiterate how cognition may be considered a tool for allostasis, specifically providing means for an organism to cope with spatially distributed resources by means of neuronal networks that can predict space and select appropriate behaviour for navigation and obstacle negotiation. Different organisms have evolved different means to take advantage of space, but depending on size and whether they are predators or herbivores, animals display cognitive abilities across a spectrum. The bacterium *E. Coli* can be contrasted with the nematode worm *C. Elegans*. The bacterium relies on molecular motors to orient itself and move in nutrition gradients, and thus can support recognition via molecular sensors, but have no behavioural choices. *C. Elegans* is multicellular, and does have a nervous system. It can recognize different types of bacteria as prey, can develop preferences for some bacteria over others, and its body supports orientation along chemical gradients. *C. Elegans* has been observed to use at least two types of locomotion. Organisms called sea squirts have life phases that include first cognition and movement in space, followed by a stationary phase, where the nervous system is dismantled, and cognition is no longer necessary. Hence *space* may be recognized as a contour of cognition, along with *recognition and choice*. Coping with space, particularly for land animals often requires distinguishing between *places*, as well as paths between a nest an a foraging place. Navigation between places requires *learning* and prediction of sequences. Learning and prediction are complementary, but prediction is directly necessary for allostasis. Hence *prediction* of sensory impression in relation to spatial navigation may be recognized as a second contour of cognition. For mammals, playing back navigation and foraging sequences, with the aid of networks that support grid representation of places, afford inferences of possible shortcuts and paths between known foraging sites. Considered as a simple form of reflection, directed and free association may also be considered a contour of cognition.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Albrecht, D., Quäschling, U., Zippel, U., & Davidowa, H. (1996). Effects of dopamine on neurons of the lateral geniculate nucleus: An iontophoretic study. *Synapse*, *23*.

Alcaro, A., Huber, R., & Panksepp, J. (2007). Behavioral functions of the mesolimbic dopaminergic system: An affective neuroethological perspective. *Brain Research Reviews*, *56*, 283–321.

Alivisatos, B., & Petrides, M. (1997). Functional activation of the human brain during mental rotation. *Neuropsychologia*, *35*, 111–118.

Arnulfo, G., Wang, S. H., Myrov, V., Toselli, B., Hirvonen, J., Fato, M., … Palva, J. (2020). Long-range phase synchronization of high-frequency oscillations in human cortex. *Nature communications*, *11*(1), 1–15.

Ashby, W. R. (1952). *Design for a brain*. London: ChapmanHall.

Baca, S. M., Marin-Burgin, A., Wagenaar, D. A., & Kristan, W. B. (2008). Widespread inhibition proportional to excitation controls the gain of a leech behavioral circuit. *Neuron*, *57*, 276–289.

Balkenius, C., Johansson, B., & Tjøstheim, T. A. (2020). Ikaros: A framework for controlling robots with system-level brain models. *International Journal of Advanced Robotic Systems*, *17*.

Balkenius, C., Tjøstheim, T. A., Johansson, B., & Gärdenfors, P. (2018). From focused thought to reveries: a memory system for a conscious robot. *Frontiers in Robotics and AI*, 29.

Balkenius, C., Tjøstheim, T. A., Johansson, B., Wallin, A., & Gärdenfors, P. (2020). The missing link between memory and reinforcement learning. *Frontiers in Psychology*, *11*, 3446.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T. P., Mirowski, P. W., … Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature*, *557*, 429–433.

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, *14*(2), 171–185.

Barnes, J. (1982). *The presocratic philosophers*. London New York: Routledge Taylor Francis Group.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*, 419–429.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Beaulieu, J. M., & Gainetdinov, R. R. (2011). The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacological Reviews*, *63*, 182–217.

Beilock, S. L., Bertenthal, B. I., Mccoy, A. M., & Carr, T. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review*, *11*, 373–379.

Bench, S. W., & Lench, H. C. (2019). Boredom as a seeking state: Boredom prompts the pursuit of novel (even negative) experiences. *Emotion*, *19*, 242–254.

Berridge, K. C. (2006). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*, 391–431.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*, 309–369.

Berridge, M. J. (2013). Dysregulation of neural calcium signaling in alzheimer disease, bipolar disorder and schizophrenia. *Prion*, *7*(1), 2–13.

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015). Burnout–depression overlap: A review. *Clinical psychology review*, *36*, 28–41.

Bird, C. D., & Emery, N. J. (2009). Rooks use stones to raise the water level to reach a floating worm. *Current Biology*, *19*, 1410–1414.

Blakemore, C., Carpenter, R. H. S., & Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, *228*, 37–39.

Blue, M. E., & Parnavelas, J. G. (1983). The formation and maturation of synapses in the visual cortex of the rat. ii. quantitative analysis. *Journal of neurocytology*, *12*(4), 697–712.

Bogler, C., Bode, S., & Haynes, J.-D. (2011). Decoding successive computational stages of saliency processing. *Current Biology*, *21*, 1667–1671.

Borg-Graham, L. J., Monier, C., & Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, *393*, 369–373.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford university press.

Botvinick, M. M., & Braver, T. S. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual review of psychology*, *66*, 83–113.

Boyle, J. H., Berri, S., Tassieri, M., Hope, I. A., & Cohen, N. (2011). Gait modulation in c. elegans: It's not a choice, it's a reflex! *Frontiers in Behavioral Neuroscience*, *5*.

Braun, C. M., & Leblanc-Sirois, Y. (2016). Beyond the synapse and hebb's rule: Is the rest of the neuron more important for psychiatric disorders? *Current Psychiatry Reviews*, *12*, 266–277.

Brillouin, L. (1956). *Science and information theory*. New York: Academic Press.

Bruineberg, J. P., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417–2444.

Bruineberg, J. P., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the new York Academy of Sciences*, *1124*(1), 1–38.

Burgess, C. R., Livneh, Y., Ramesh, R. N., & Andermann, M. L. (2018). Gating of visual processing by physiological need. *Current Opinion in Neurobiology*, *49*, 16–23.

Burgess, C. R., Ramesh, R. N., Sugden, A. U., Levandowski, K. M., Minnig, M. A., Fenselau, H., … Andermann, M. L. (2016). Hunger-dependent enhancement of food cue responses in mouse postrhinal cortex and lateral amygdala. *Neuron*, *91*, 1154–1169.

Bush, D., Barry, C., Manson, D., & Burgess, N. (2015). Using grid cells for navigation. *Neuron*, *87*, 507–520.

Butz, M. V. (2016). Toward a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology*, *7*.

Buzsáki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press.

Buzsáki, G., & Tingley, D. (2018). Space and Time: The Hippocampus as a Sequence Generator. *Trends in Cognitive Sciences*, *22*(10), 853–869.

Caduff, D., & Timpf, S. (2007). On the assessment of landmark salience for human navigation. *Cognitive Processing*, *9*, 249–267.

Camilleri, M., Atanasova, E., Carlson, P. J., Ahmad, U., Kim, H. J., Viramontes, B. E., … Urrutia, R. (2002). Serotonin-transporter polymorphism pharmacogenetics in diarrhea-predominant irritable bowel syndrome. *Gastroenterology*, *123*(2), 425–432.

Cannon, W. B. (1932). *The wisdom of the body*. London: Kegan Paul and Co., Ltd.

Cara, G. L., Bettini, M., & Ursino, M. (2003). Extraction of salient contours in primary visual cortex: a neural network model based on physiological knowledge. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *3*, 2242–2245 Vol.3.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*, *17*, 8621–8644.

Carmichael, S. T., & Price, J. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, *363*.

Carmichael, S. T., & Price, J. L. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *346*.

Carriere, J. S. A., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and Cognition*, *17*, 835–847.

Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion*, *4 1*, 3–22.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*, 105–110.

Chandrasekar, R. (2013). Alcohol and nmda receptor: current research and future direction. *Frontiers in molecular neuroscience*, *6*, 14.

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20*(2), 96–106.

Changizi, M. (2010). *The vision revolution: How the latest research overturns everything we thought we knew about human vision*. Benbella books.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical population biology*, *9 2*, 129–36.

Choi, D. W. (1985). Glutamate neurotoxicity in cortical cell culture is calcium dependent. *Neuroscience Letters*, *58*, 293–297.

Choi, D. W. (1992). Excitotoxic cell death. *Journal of neurobiology*, *23 9*, 1261–76.

Chow, J. J., Nickell, J. R., Darna, M., & Beckmann, J. S. (2016). Toward isolating the role of dopamine in the acquisition of incentive salience attribution. *Neuropharmacology*, *109*, 320–331.

Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, *58*, 345.

Cohen, I. R. (1992). The cognitive principle challenges clonal selection. *Immunology today*, *13*(11), 441–444.

Cohen, I. R. (2000). *Tending adam's garden: evolving the cognitive immune self*. Elsevier.

Colaizzi, J. M., Flagel, S. B., Joyner, M. A., Gearhardt, A. N., Stewart, J. L., & Paulus, M. P. (2020). Mapping sign-tracking and goal-tracking onto human behaviors. *Neuroscience & Biobehavioral Reviews*, *111*, 84–94.

Cole, H. L., M., & McDermott, R. P. (1997). Concepts of ecological validity: Their differing implications for comparative cognitive research. In *Mind, culture, and activity: Seminal papers from the laboratory of comparative human cognition* (pp. 49–56).

Connolly, N., & Prehn, J. H. (2015). The metabolic response to excitotoxicity–lessons from single-cell imaging. *Journal of bioenergetics and biomembranes*, *47*(1), 75–88.

Costa, V. D., Kakalios, L. C., & Averbeck, B. B. (2016). Blocking serotonin but not dopamine reuptake alters neural processing during perceptual decision making. *Behavioral neuroscience*, *130 5*, 461–8.

Coughlan, G., Laczó, J., Hort, J., Minihane, A.-M., & Hornberger, M. (2018). Spatial navigation deficits—overlooked cognitive marker for preclinical alzheimer disease? *Nature Reviews Neurology*, *14*(8), 496–506.

Craig, A. D. (2002). How do you feel? interoception: the sense of the physiological condition of the body. *Nature reviews neuroscience*, *3*(8), 655–666.

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge University Press.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*, 189–195.

Crosson, F. J., & Sayre, K. M. (1968). Philosophy and cybernetics. New York: Simon and Schuster.

Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. WW Norton & Company.

De Vadder, F., Grasset, E., Holm, L. M., Karsenty, G., Macpherson, A. J., Olofsson, L. E., & Bäckhed, F. (2018). Gut microbiota regulates maturation of the adult enteric nervous system via enteric serotonin networks. *Proceedings of the National Academy of Sciences*, *115*(25), 6458–6463.

Dietterich, T. G. (1995). Overfitting and undercomputing in machine learning. *ACM Comput. Surv.*, *27*, 326–327.

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences*, *105*(15), 5915–5920.

Dölen, G., Darvishzadeh, A., Huang, K. W., & Malenka, R. C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature*, *501*(7466), 179–184.

Donovan, M. H., & Tecott, L. H. (2013). Serotonin and the regulation of mammalian energy balance. *Frontiers in neuroscience*, *7*, 36.

Dudek, S. M., & Bear, M. F. (1992). Homosynaptic long-term depression in area ca1 of hippocampus and effects of n-methyl-d-aspartate receptor blockade. *Proceedings of the National Academy of Sciences*, *89*, 4363–4367.

Dunn, L. T., & Everitt, B. J. (1988). Double dissociations of the effects of amygdala and insular cortex lesions on conditioned taste aversion, passive avoidance, and neophobia in the rat using the excitotoxin ibotenic acid. *Behavioral neu-*

roscience, *102 1*, 3–23.

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, *7*(5), 482–495.

Ebert, U., & Ostwald, J. (1992). Serotonin modulates auditory information processing in the cochlear nucleus of the rat. *Neuroscience Letters*, *145*, 51–54.

Engel, A. K., Kreiter, A. K., König, P., & Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proceedings of the National Academy of Sciences*, *88 14*, 6048–52.

Erlich, J. C., Bush, D. E. A., & LeDoux, J. E. (2012). The role of the lateral amygdala in the retrieval and maintenance of fear-memories formed by repeated probabilistic reinforcement. *Frontiers in Behavioral Neuroscience*, *6*.

Esghaei, M., Treue, S., & Vidyasagar, T. R. (2022). Dynamic coupling of oscillatory neural activity and its roles in visual attention. *Trends in Neurosciences*.

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, *21*, 80–92.

Fiebelkorn, I. C., & Kastner, S. (2019). A rhythmic theory of attention. *Trends in cognitive sciences*, *23*(2), 87–101.

Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, *86*, 229–242.

French, M. (2018). Minimum principles: Optimization in the fabric of the universe. In *Fundamentals of optimization: Methods, minimum principles, and applications for making things better* (pp. 55–70). Cham: Springer International Publishing.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of physiology-Paris*, *100*(1–3), 70–87.

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, *6*(4), 187–214.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area v4 of the macaque monkey. *Journal of neurophysiology*, *76 4*, 2718–39.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., … Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, 1–6.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. J. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, *6*.

Geary, D. C. (2018). Efficiency of mitochondrial functioning as the fundamental biological mechanism of general intelligence (g). *Psychological Review*, *125*(6), 1028.

Gerrans, P. (2014). *The measure of madness: Philosophy of mind, cognitive neuroscience, and delusional thought.* MIT Press.

Geurts, L. S., Cooke, J. R. H., van Bergen, R. S., & Jehee, J. F. M. (2021). Subjective confidence reflects representation of bayesian probability in cortex. *bioRxiv*.

Gevins, A. S., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral cortex*, *10 9*, 829–39.

Gidlöf, K., Anikin, A., Lingonblad, M., & Wallin, A. (2017). Looking is buying. how visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite*, *116*, 29–38.

Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, *6*(1).

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15647–15654.

Glinoer, D. (1997). The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocrine reviews*, *18 3*, 404–33.

Goetz, T., Frenzel, A. C., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation and Emotion*, *38*, 401–419.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*, 535–74.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20–25.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M. C., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*, 585–593.

Govindaiah, G., & Cox, C. L. (2005). Excitatory actions of dopamine via d1-like receptors in the rat lateral geniculate nucleus. *Journal of neurophysiology*, *94 6*, 3708–18.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, *86 5*, 1698–702.

Gray, J., & McNaughton, N. (2000). *The neuropsychology of anxiety.* New York: Oxford University Press.

Greer-Phillips, S. E., Alexandre, G., Taylor, B. L., & Zhulin, I. B. (2003). Aer and tsr guide escherichia coli in spatial gradients of oxidizable substrates. *Microbiology*, *149 Pt 9*, 2661–7.

Greicius, M. D., Krasnow, B. D., Reiss, A. L., & Menon, V. (2002). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*, 253–258.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., … Bassett, D. S. (2015). Controllability of structural brain networks. *Nature Communications*, *6*.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, Massachusetts: MIT press.

Ha, G. E., & Cheong, E. (2017). Spike frequency adaptation in neurons of the central nervous system. *Experimental Neurobiology*, *26*, 179–185.

Haenisch, B., & Bönisch, H. (2011). Depression and antidepressants: insights from knockout of dopamine, serotonin or noradrenaline re-uptake transporters. *Pharmacology & therapeutics*, *129 3*, 352–68.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of experimental psychology*, *58*, 321–34.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*, 1726–1731.

Hebb, D. (1949). Organization of behavior. new york: Wiley. *J. Clin. Psychol*, *6*(3), 335–307.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual neuroscience*, *9 2*, 181–97.

Heinrich, B., & Bugnyar, T. (2005). Testing problem solving in ravens: String-pulling to reach food. *Ethology*, *111*, 962–976.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, *117*.

Hoffmann, J. (2003). Anticipatory behavioral control. In *Anticipatory behavior in adaptive learning systems* (pp. 44–65). Springer.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109 4*, 679–709.

Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Computation*, *9*, 1001–1013.

Holtz, K., Pané-Farré, C. A., Wendt, J., Lotze, M., & Hamm, A. O. (2012). Brain activation during anticipation of interoceptive threat. *NeuroImage*, *61*, 857–865.

Holyoak, K. J., & Morrison, R. G. (2012). Thinking and reasoning: A reader's guide. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning.* Oxford University Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (tec): a framework for perception and action planning. *The Behavioral and brain sciences*, *24 5*, 849–78; discussion 878–937.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554–2558.

Huang, Z., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M. F., … Tonegawa, S. (1999). Bdnf regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell*, *98*, 739–755.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*.

Hurley, L. M., & Pollak, G. D. (1999). Serotonin differentially modulates responses to tones and frequency-modulated sweeps in the inferior colliculus. *The Journal of Neuroscience*, *19*, 8071–8082.

Hyrc, K. L., Handran, S. D., Rothman, S. M., & Goldberg, M. P. (1997). Ionized intracellular calcium concentration predicts excitotoxic neuronal death: Observations with low-affinity fluorescent calcium indicators. *The Journal of Neuroscience*, *17*, 6669–6677.

Ikonomidou, C., Bittigau, P., Koch, C., Genz, K., Hoerster, F., Felderhoff-Mueser, U., … Olney, J. W. (2001). Neurotransmitters and apoptosis in the developing brain. *Biochemical pharmacology*, *62*(4), 401–405.

Ikonomidou, C., Bosch, F., Miksa, M., Bittigau, P., Vöckler, J., Dikranian, K., … Olney, J. W. (1999). Blockade of nmda receptors and apoptotic neurodegeneration in the developing brain. *Science*, *283 5398*, 70–4.

Ito, M. (2013). Error detection and representation in the olivo-cerebellar system. *Frontiers in Neural Circuits*, *7*(February), 1–8.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE transactions on neural networks*, *14 6*, 1569–72.

Jacob, S. N., & Nienborg, H. (2018). Monoaminergic neuromodulation of sensory processing. *Frontiers in Neural Circuits*, *12*.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079.

Jonsson, B., Kulaksiz, Y. C., & Lithner, J. (2016). Creative and algorithmic mathematical reasoning: effects of transfer-appropriate processing and effortful struggle. *International Journal of Mathematical Education in Science and Technology*, *47*, 1206–1225.

Kabadayi, C., & Osvath, M. (2017). Ravens parallel great apes in flexible planning for tool-use and bartering. *Science*, *357*(6347), 202–204.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American journal of Psychiatry*, *160*(1), 13–23.

Kasamatsu, T., & Heggelund, P. (1982). Single cell responses in cat visual cortex to visual stimulation during iontophoresis of noradrenaline. *Experimental Brain Research*, *45*, 317–327.

Keijzer, F., & Arnellos, A. (2017). The animal sensorimotor organization: a challenge for the environmental complexity thesis. *Biology & philosophy*, *32*(3), 421–441.

Keil, A., Müller, M. M., Ray, W. J., Gruber, T., & Elbert, T. (1999). Human gamma band activity and perception of a gestalt. *Journal of Neuroscience*, *19*(16), 7152–7161.

Kennedy, P. J., & Shapiro, M. L. (2009). Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. *Proceedings of the National Academy of Sciences*, *106*, 10805–10810.

Kesner, R. P., & Creem-Regehr, S. H. (2013). Parietal contributions to spatial cognition. In D. Waller & L. Nadel (Eds.), *Handbook of spatial cognition* (pp. 35–63). Washington: American Psychological Association.

Kirkwood, A., Lee, H.-K., & Bear, M. F. (1995). Co-regulation of long-term potentiation and experience-dependent synaptic plasticity in visual cortex by age and experience. *Nature*, *375*(6529), 328–331.

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., … Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, *1*(5), 1–37.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of neurophysiology*, *71 3*, 856–67.

Kolb, B. E., Pedersen, B., Ballermann, M. A., Gibb, R. L., & Whishaw, I. Q. (1999). Embryonic and postnatal injections of bromodeoxyuridine produce age-dependent morphological and behavioral abnormalities. *The Journal of Neuroscience*, *19*, 2337–2346.

Konczak, J., Borutta, M., Topka, H., & Dichgans, J. (2004). The development of goal-directed reaching in infants: hand trajectory formation and joint torque control. *Experimental Brain Research*, *106*, 156–168.

Kornblith, H. (2012). *On reflection*. Oxford University Press.

Kotikova, E. A., Raikova, O. I., Reuter, M., & Gustafsson, M. K. (2005). Rotifer nervous system visualized by fmrfamide and 5-ht immunocytochemistry and confocal laser scanning microscopy. *Hydrobiologia*, *546*(1), 239–248.

Kucyi, A., & Davis, K. D. (2014). Dynamic functional connectivity of the default mode network tracks daydreaming. *NeuroImage*, *100*, 471–480.

Köhler, W. (1925). *The mentality of apes*. Harcourt, Brace.

Lane, N. (2005). *Power, sex, suicide: Mitochondria and the meaning of life*. London: Oxford University Press.

Lendvai, B., & Vizi, E. S. (2008). Nonsynaptic chemical transmission through nicotinic acetylcholine receptors. *Physiological reviews*, *88 2*, 333–49.

Lewis, B. L., & O'Donnell, P. (2000). Ventral tegmental area afferents to the prefrontal cortex maintain membrane potential 'up' states in pyramidal neurons via d(1) dopamine receptors. *Cerebral cortex*, *10 12*, 1168–75.

Lewis, J. (2006). Cortical networks related to human use of tools. *The Neuroscientist*, *12*, 211–231.

Li, W., Mai, X., & Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Frontiers in Human Neuroscience*, *8*.

Lin, T.-W., & Kuo, Y.-M. (2013). Exercise benefits brain function: The monoamine connection. *Brain Sciences*, *3*, 39–53.

Lissek, S. M., Rabin, S. J., Heller, R., Lukenbaugh, D., Geraci, M. F., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *The American journal of psychiatry*, *167 1*, 47–55.

Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2011). Controllability of complex networks. *Nature*, *473*(7346), 167–173.

Livneh, Y., Ramesh, R. N., Burgess, C. R., Levandowski, K. M., c. Madara, J., Fenselau, H., … Andermann, M. L. (2017). Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature*, *546*, 611–616.

Lorenz, K. (1977). Behind the mirror: A search for a natural history of human knowledge. Harcourt Brace Jovanovich.

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: Divisive normalization in parietal cortex. *The Journal of Neuroscience*, *31*, 10627–10639.

Lovero, K. L., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2009). Anterior insular cortex anticipates impending stimulus significance. *NeuroImage*, *45*, 976–983.

Mansur, R. B., Lee, Y., McIntyre, R. S., & Brietzke, E. (2020). What is bipolar disorder? a disease model of dysregulated energy expenditure. *Neuroscience & Biobehavioral Reviews*, *113*, 529–545.

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience*, *6*.

Martin-Ordas, G., & Call, J. (2009). Assessing generalization within and between trap tasks in the great apes. *International Journal of Comparative Psychology*, *22*(1).

Martin-Ordas, G., Call, J., & Colmenares, F. (2008). Tubes, tables and traps: great apes solve two functionally equivalent trap tasks but show no evidence of transfer across tasks. *Animal cognition*, *11*(3), 423–430.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*, 837–841.

Mcafee, S. S., Liu, Y., Sillitoe, R. V., & Heck, D. H. (2022). Cerebellar Coordination of Neuronal Communication in Cerebral Cortex. , *15*(January).

McCorry, L. K. (2007). Physiology of the autonomic nervous system. *American journal of pharmaceutical education*, *71*(4).

Merrer, J. L., Becker, J. A. J., Befort, K., & Kieffer, B. L. (2009). Reward processing by the opioid system in the brain. *Physiological reviews*, *89 4*, 1379–412.

Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, *121 (Pt 6)*, 1013–52.

Milivojevic, B., Varadinov, M., Grabovetsky, A. V., Collin, S. H., & Doeller, C. F. (2016). Coding of event nodes and narrative context in the hippocampus. *The Journal of Neuroscience*, *36*, 12412–12424.

Miller, L. (2012). Stalking: Patterns, motives, and intervention strategies. *Aggression and violent behavior*, *17*(6), 495–506.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785.

Mirshahi, T., & Woodward, J. J. (1995). Ethanol sensitivity of heteromeric nmda receptors: Effects of subunit assembly, glycine and nmdar1 mg2+-insensitive mutants. *Neuropharmacology*, *34*, 347–355.

Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world. *PLoS ONE*, *13*.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, *6*, 414–417.

Mogg, K., Bradley, B. P., Hyare, H., & Lee, S. C. (1998). Selective attention to food-related stimuli in hunger: are attentional biases specific to emotional and psychopathological states, or are they also found in normal drive states? *Behaviour research and therapy*, *36 2*, 227–37.

Nakazawa, K., McHugh, T. J., Wilson, M. A., & Tonegawa, S. (2004). Nmda receptors, place cells and hippocampal spatial memory. *Nature Reviews Neuroscience*, *5*, 361–372.

Neophytou, D., Arribas, D. M., Levy, R. B., Park, M., & Oviedo, H. V. (2021). Recurrent connectivity underlies lateralized temporal processing differences in auditory cortex. *bioRxiv*.

Neve, K. A., Seamans, J. K., & Trantham-Davidson, H. (2004). Dopamine receptor signaling. *Journal of Receptors and Signal Transduction*, *24*, 165–205.

Nieh, E. H., Weele, C. M. V., Matthews, G. A. C., Presbrey, K. N., Wichmann, R., Leppla, C. A., … Tye, K. M. (2016). Inhibitory input from the lateral hypothalamus to the ventral tegmental area disinhibits dopamine neurons and promotes behavioral activation. *Neuron*, *90*, 1286–1298.

Nongonierma, A. B., & FitzGerald, R. J. (2015). Milk proteins as a source of tryptophan-containing bioactive peptides. *Food & Function*, *6*(7), 2115–2127.

Noudoost, B., & Moore, T. (2011a). Control of visual cortical signals by prefrontal dopamine. *Nature*, *474*, 372–375.

Noudoost, B., & Moore, T. (2011b). The role of neuromodulators in selective attention. *Trends in Cognitive Sciences*, *15*, 585–591.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton, New Jersey, UK: Princeton University Press.

Olsen, S. R., Bhandawat, V., & Wilson, R. I. (2010). Divisive normalization in olfactory population codes. *Neuron*, *66*, 287–299.

Osvath, M., & Osvath, H. (2008). Chimpanzee (Pan troglodytes) and orangutan (Pongo abelii) forethought: Self-control and pre-experience in the face of future tool use. *Animal Cognition*, *11*(4), 661–674.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, *51*, 78–109.

Pagotto, U., Marsicano, G., Cota, D., Lutz, B., & Pasquali, R. (2006). The emerging role of the endocannabinoid system in endocrine regulation and energy balance. *Endocrine reviews*, *27 1*, 73–100.

Pape, L., Oddo, C. M., Controzzi, M., Cipriani, C., Förster, A., Carrozza, M. C., & Schmidhuber, J. (2012). Learning tactile skills through curious exploration. *Frontiers in Neurorobotics*, *6*.

Papini, M. R. (2002). Pattern and process in the evolution of learning. *Psychological review*, *109*(1), 186.

Passingham, R. E., & Wise, S. P. (2012). *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight*. Oxford University Press.

Patai, E. Z., & Spiers, H. J. (2021). The versatile wayfinder: prefrontal contributions to spatial navigation. *Trends in cognitive sciences*, *25*(6), 520–533.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 488–489.

Pauling, L. (1940). *The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structural chemistry* (2nd ed ed.). London: Oxford University Press.

Peciña, S., & Berridge, K. C. (2005). Hedonic hot spot in nucleus accumbens shell: Where do μ-opioids cause increased hedonic impact of sweetness? *The Journal of Neuroscience*, *25*, 11777–11786.

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in neurobiology*, *156*, 164–188.

Petoukhov, S. V. (2016). The system-resonance approach in modeling genetic structures. *Biosystems*, *139*, 1–11.

Phillis, J. W., & Tebēcis, A. K. (1967). The responses of thalamic neurones to iontophoretically applied monoamines. *The Journal of Physiology*, *192*.

Poe, G. R., Nitz, D. A., McNaughton, B. L., & Barnes, C. A. (2000). Experience-dependent phase-reversal of hippocampal neuron firing during rem sleep. *Brain Research*, *855*, 176–180.

Poulter, S., Hartley, T., & Lever, C. (2018). The neurobiology of mammalian navigation. *Current Biology*, *28*(17), R1023–R1042.

Poulter, S., Lee, S. A., Dachtler, J., Wills, T. J., & Lever, C. (2021). Vector trace cells in the subiculum of the hippocampal formation. *Nature neuroscience*, *24*(2), 266–275.

Pratt, W. E., & Kelley, A. E. (2004). Nucleus accumbens acetylcholine regulates appetitive learning and motivation for food via activation of muscarinic receptors. *Behavioral neuroscience*, *118*(4), 730.

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., & King, A. J. (2011). Contrast gain control in auditory cortex. *Neuron*, *70*, 1178–1191.

Raffaelli, Q., Mills, C., & Christoff, K. (2017). The knowns and unknowns of boredom: a review of the literature. *Experimental Brain Research*, *236*, 2451–2462.

Ramos, B. P., & Arnsten, A. F. (2007). Adrenergic pharmacology and cognition: focus on the prefrontal cortex. *Pharmacology & therapeutics*, *113*(3), 523–536.

Rao, R. P. N., & Ballard, D. H. (1998). Development of localized oriented receptive fields by learning a translation-invariant code for natural images. *Network*, *9 2*, 219–34.

Raz, L., Knoefel, J., & Bhaskar, K. (2016). The neuropathology and cerebrovascular mechanisms of dementia. *Journal of Cerebral Blood Flow & Metabolism*, *36*(1), 172–186.

Reas, C., & Fry, B. (2007). Processing: a programming handbook for visual designers and artists. MIT Press.

Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, *17*(3), 147–159.

Redish, A. D., & Touretzky, D. S. (1998). The role of the hippocampus in solving the morris water maze. *Neural Computation*, *10*, 73–111.

Reznikova, Z., Levenets, J., Panteleeva, S., & Ryabko, B. Y. (2017). Studying hunting behaviour in the striped field mouse using data compression. *acta ethologica*, *20*, 165–173.

Ritz, T., Wiltschko, R., Hore, P., Rodgers, C. T., Stapput, K., Thalau, P., … Wiltschko, W. (2009). Magnetic compass of birds is based on a molecule with optimal directional sensitivity. *Biophysical journal*, *96*(8), 3451–3457.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., … Vanduffel, W. (2012). Toward a unified theory of visual area v4. *Neuron*, *74*(1), 12–29.

Rogawski, M. A., & Aghajanian, G. K. (1980). Modulation of lateral geniculate neurone excitability by noradrenaline microiontophoresis or locus coeruleus stimulation. *Nature*, *287*, 731–734.

Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *The Journal of Neuroscience*, *23*, 339–348.

Rushworth, M. F. S. (2008). Intention, choice, and the medial frontal cortex. *Annals of the New York Academy of Sciences*, *1124*.

Ryan, K., Lu, Z., & Meinertzhagen, I. A. (2016). The cns connectome of a tadpole larva of ciona intestinalis (l.) highlights sidedness in the brain of a chordate sibling. *eLife*, *5*.

Salomon, G. (1992). Transfer of learning. In *International encyclopedia of education 2* (pp. 6452–6457).

Satoh, M., & Minami, M. (1995). Molecular pharmacology of the opioid receptors. *Pharmacology & therapeutics*, *68*(3), 343–364.

Sattler, R., Xiong, Z.-G., Lu, W. Y., Hafner, M., Macdonald, J. F., & Tymianski, M. (1999). Specific coupling of nmda receptor activation to nitric oxide neurotoxicity by psd-95 protein. *Science*, *284 5421*, 1845–8.

Sayre, K. M. (1976). Cybernetics and the philosophy of mind. Routledge and Kegan Paul.

Serino, S., Cipresso, P., Morganti, F., & Riva, G. (2014). The role of egocentric and allocentric abilities in alzheimer's disease: a systematic review. *Ageing research reviews*, *16*, 32–44.

Shadlen, M., & Kandel, E. (2021a). The hypothalamus: Autonomic, hormonal, and behavioral control of survival. In *Principles of neural science* (6th ed.). New York, NY: McGraw Hill.

Shadlen, M., & Kandel, E. (2021b). Synaptic integration in the central nervous system. In *Principles of neural science* (6th ed.). New York, NY: McGraw Hill.

Shamash, P., Olesen, S. F., Iordanidou, P., Campagner, D., Banerjee, N., & Branco, T. (2021). Mice learn multi-step routes by memorizing subgoal locations. *Nature Neuroscience*, *24*(9), 1270–1279.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*(7), 379–423, 623–656.

Shettleworth, S. J. (1989). Animals foraging in the lab: Problems and promises. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 81–87.

Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K. Telesford, Alfred B. Yu, Ari E. Kahn, … Danielle S. Bassett (2015). Controllability of structural brain networks. *Nature Communications*, *6*, 8414.

Shinkai, Y., Yamamoto, Y., Fujiwara, M., Tabata, T., Murayama, T., Hirotsu, T., … Ishihara, T. (2011). Behavioral choice between conflicting alternatives is regulated by a receptor guanylyl cyclase, gcy-28, and a receptor tyrosine kinase, scd-2, in aia interneurons of caenorhabditis elegans. *The Journal of Neuroscience*, *31*, 3007–3015.

Shtonda, B. B., & Avery, L. (2006). Dietary choice behavior in caenorhabditis elegans. *Journal of Experimental Biology*, *209*, 102–89.

Silva, M. F., Harvey, B. M., Jorge, L., Canário, N., Machado, F., Soares, M., … Castelo-Branco, M. (2021). Simultaneous changes in visual acuity, cortical population receptive field size, visual field map size, and retinal thickness in healthy human aging. *Brain Structure & Function*, *226*, 2839–2853.

Silverman, M. C., Nies, D., Jung, B., & Sukhatme, G. S. (2002). Staying alive: a docking station for autonomous robot recharging. *Proceedings 2002 IEEE International Conference on Robotics and Automation*, *1*, 1050–1055.

Smallwood, J., Brown, K. S., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Research*, *1428*, 60–70.

Smith, K. L., Lee, C. L., & Swann, J. W. (1998). Local circuit abnormalities in chronically epileptic rats after intrahippocampal tetanus toxin injection in infancy. *Journal of neurophysiology*, *79 1*, 106–16.

Smith, K. S., & Berridge, K. C. (2005). The ventral pallidum and hedonic reward: Neurochemical maps of sucrose "liking" and food intake. *The Journal of Neuroscience*, *25*, 8637–8649.

Soh, N. L., & Walter, G. (2011). Tryptophan and depression: can diet alone be the answer? *Acta Neuropsychiatrica*, *23*(1), 3–11.

Stanika, R. I., Pivovarova, N. B., Brantner, C. A., Watts, C. A., Winters, C., & Andrews, S. B. (2009). Coupling diverse routes of calcium entry to mitochondrial dysfunction and glutamate excitotoxicity. *Proceedings of the National Academy of Sciences*, *106*, 9854–9859.

Stefanatos, R., & Sanz, A. (2018). The role of mitochondrial ros in the aging brain. *FEBS letters*, *592*(5), 743–758.

Stephens, A., & Tjøstheim, T. A. (2020). The cognitive philosophy of reflection. *Erkenntnis*, 1–24.

Takagi, S., Nishiura, Y., Nakagaki, T., Ueda, T., & Ueda, K.-I. (2007). Indecisive behavior of amoeba crossing an environmental barrier. In *Topological aspects of critical systems and networks: (with cd-rom)* (pp. 86–93). World Scientific.

Tärning, B., Lee, Y. J., Andersson, R., Månsson, K., Gulz, A., & Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *Journal of the Learning Sciences*, *29*(4–5), 511–549.

Tecott, L. H. (2007). Serotonin and the Orchestration of Energy Balance. *Cell Metabolism*, *6*(5), 352–361.

Terrace, H. S. (1963). Discrimination learning with and without "errors". *Journal of the experimental analysis of behavior*, *6*, 1–27.

Tiernan, B., Tracey, R., & Shannon, C. (2014). Paranoia and self-concepts in psychosis: a systematic review of the literature. *Psychiatry research*, *216*(3), 303–313.

Tjøstheim, T. A., & Balkenius, C. (2019). Cumulative inhibition in neural networks. *Cognitive Processing*, *20*(1), 87–102.

Tjøstheim, T. A., Johansson, B., & Balkenius, C. (2019). A computational model of trust-, pupil-, and motivation dynamics. In *Proceedings of the 7th international conference on human-agent interaction* (pp. 179–185).

Tjøstheim, T. A., Johansson, B., & Balkenius, C. (2021). Direct approach or detour: A comparative model of inhibition and neural ensemble size in behavior selection. *Frontiers in Systems Neuroscience*, *15*.

Tjøstheim, T. A., & Stephens, A. (2021). Intelligence as accurate prediction. *Review of Philosophy and Psychology*, 1–25.

Vallverd'u, J., Castro, O., Mayne, R., Talanov, M., Levin, M., Baluška, F., … Adamatzky, A. (2018). Slime mould: the fundamental mechanisms of biological cognition. *Biosystems*, *165*, 57–70.

van Bergen, R. S., & Jehee, J. F. M. (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *The Journal of Neuroscience*, *39*, 8164–8176.

van Lawick-Goodall, J. (1971). Tool-using in primates and other vertebrates. *Advances in The Study of Behavior*, *3*, 195–249.

van Steenbergen, H., Eikemo, M., & Leknes, S. (2019). The role of the opioid system in decision making and cognitive control: A review. *Cognitive, Affective & Behavioral Neuroscience*, *19*, 435–458.

Visalberghi, E. (1990). Tool use in cebus. *Folia primatologica; international journal of primatology*, *54 3–4*, 146–54.

von Uexküll, J. (1934). Streifzge durch die unwelten von tieren und menschen (a stroll through the worlds of animal and men). *Instinctive behavior*, 5–80.

Wagner, M. J., & Smith, M. A. (2008). Shared internal models for feedforward and feedback control. *The Journal of Neuroscience*, *28*, 10663–10673.

Wallace, R., & Wallace, R. (2017). *Computational psychiatry*. Springer.

Wallis, J. D., & Rushworth, M. F. (2014). Integrating benefits and costs in decision making. In *Neuroeconomics* (pp. 411–433). Elsevier.

Wang, D. (1995). Emergent synchrony in locally coupled neural oscillators. *IEEE transactions on neural networks*, *6 4*, 941–8.

Wang, D., & Terman, D. H. (1997). Image segmentation based on oscillatory correlation. *Neural Computation*, *9*, 805–836.

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*(3), 279–292.

Weber, S. N., & Sprekeler, H. (2018). Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *eLife*, *7*.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode caenorhabditis elegans. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *314 1165*, 1–340.

Whitehead, A. N. (1929). *Process and reality*. New York: Macmillan.

Whittington, M. A., & Jefferys, J. G. R. (1994). Epileptic activity outlasts disinhibition after intrahippocampal tetanus toxin in the rat. *The Journal of Physiology*, *481*.

Wiener, N. (1948). Cybernetics: Or control and communication in the animal and the machine. MIT Press.

Williams, A. M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research quarterly for exercise and sport*, *69 2*, 111–28.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*, 483–494.

Woodin, G., & Winter, B. (2018). Placing abstract concepts in space: Quantity, time and emotional valence. *Frontiers in Psychology*, *9*.

Worthy, S. E., Rojas, G. L. M., Taylor, C. J., & Glater, E. E. (2018). Identification of odor blend used by caenorhabditis elegans for pathogen recognition. *Chemical Senses*, *43*, 169–180.

Wright, J. S., & Panksepp, J. (2012, jan). An Evolutionary framework to understand foraging, wanting, and Desire: The Neuropsychology of the SEEKING system. *Neuropsychoanalysis*, *14*(1), 5–39.

Xiao, Z., Lin, K. K., & Fellous, J.-M. (2020). Conjunctive reward–place coding properties of dorsal distal ca1 hippocampus cells. *Biological Cybernetics*, *114*, 285–301.

Xiong, W., & Chen, W. R. (2002). Dynamic gating of spike propagation in the mitral cell lateral dendrites. *Neuron*, *34*, 115–126.

Yakobi, O., & Danckert, J. (2021). Boredom proneness is associated with noisy decision-making, not risk-taking. *Experimental Brain Research*, 1–19.

Yang, G. R., Murray, J. D., & Wang, X.-J. (2016). A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nature Communications*, *7*.

Yoon, W., Won, S., Ryu, B., & Gwag, B. (2003). Blockade of ionotropic glutamate receptors produces neuronal apoptosis through the bax-cytochrome c-caspase pathway: the causative role of ca2+ deficiency. *Journal of neurochemistry*, *85*(2), 525–533.

Yoshida, M., Takayanagi, Y., Inoue, K., Kimura, T., Young, L. J., Onaka, T., & Nishimori, K. (2009). Evidence that oxytocin exerts anxiolytic effects via oxytocin receptor expressed in serotonergic neurons in mice. *Journal of Neuroscience*, *29*(7), 2259–2271.

Zach, R. (1979). Shell dropping: Decision-making and optimal foraging in northwestern crows. *Behaviour*, *68*, 106–117.

Zaldivar, D., Goense, J. B. M., Lowe, S. C., Logothetis, N. K., & Panzeri, S. (2018). Dopamine is signaled by mid-frequency oscillations and boosts output layers visual information in visual cortex. *Current Biology*, *28*, 224–235.e5.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22*, 775–789.

Index

Scientific publications

Summary, results, and contributions of work

Paper I: From Focused Thought to Reveries – A Memory System for a Conscious Robot

This paper presents a computational model of various kinds of memory, like working memory and episodic memory, along with an arousal system inspired by noradrenergic systems in the mammalian brain, and that is relevant for the study of awareness and different conscious states like day–dreaming and focused goal approach. In the context of foraging, this paper deals with knowledge and memory about which patterns exist in an ecological niche at which places, as well as how arousal affects what an organism may be aware of.

Result: A model of memory and arousal that can produce responses to mediate several kinds cognitive processes, including object permanence, episodic recall, and vicarious trial and error. The same model was successfully tested in a range of psychological tasks including the A–not–B task and delayed matching to sample.

Contribution: Provides a framework for studying integrated memory– and spatial processes in a systemically plausible way, associating object recognition with place perception. Support for arousal dependent free association provides a basis for studying mental behaviour associated with default network processes, awareness, and consciousness.

Paper II: Cumulative inhibition in neural networks

This paper proposes a neural network algorithm for unsupervised learning that supports multi–resolution receptive fields. Units can be stacked hierarchically to arbitrary depth to cover increasingly large areas of a sensory field. We suggest that this network provides a model for understanding the development of acuity in perception, where coarse representations precede representations with fine detail. In the context of foraging, this paper thus deals with perception and development, and what the statistical sensory building blocks of an environmental niche are.

Result: Network units adapt their weights to the statistics of the input, having a qualitatively similar acuity stabilization curve as that shown in kittens. The cumulative inhibition topology produces network weights that have progressively higher power at high spatial frequencies.

Contribution: Provides an algorithm that supports detail-differentiated statistical learning, and provides steps toward a theoretical framework for the development of perceptual acuity.

Paper III: A computational model of trust, pupil, and motivation dynamics

This paper extends an earlier published model of pupil dilation and adds nuclei and pathways that mediate familiarity based trust and valenced somatic sensation in the form of gentle and painful touch. From the perspective of foraging, this paper shows how familiarity and positive feelings are building blocks of unconditional trust, and how trust –or its absence– can be a differentiator for approach and avoidance behaviour.

Result: Simulations of faces associated with gentle or painful touch suggest that stimuli inducing negatively valenced affect, induce stronger learning than those that induce positively valenced affect. This is reflected in a larger diameter of pupil dilation for negative stimuli compared to positive, and yields an asymmetrically stronger motivation to avoid, than positive stimuli do to approach. Results also indicate that given sequential presentation of both negative and positive stimuli of equal magnitude, the negative stimulus will dominate the positive one regarding level of trust.

Contribution: A working and anatomically plausible computational model of basic familiarity based trust, applicable to real–time performance in robots.

Paper IV: The cognitive philosophy of reflection

According to Kornblith, only direct experience can add reliable knowledge, in contrast to reflection which cannot. This paper argues that a less restrictive view of reflection is commensurable with empirical results from neuroscience and cognitive science, and that reflection can add reliability through generalization, flexible association, and creativity. Following successful generalization from experience, reflection can provide reliable behavioural guides in novel situations. We therefore propose that knowledge can be usefully viewed as having a component of direct experience, as well as one based on reflection. Therefore, from the perspective of a foraging animal, reflection, thinking back on experiences and using knowledge from different sources can increase reliability of future predictions and make for better choices in novel situation.

Contribution: Shows how reflection can contribute epistemically apart from direct experience. A way to understand thinking, reflection that is commensurate with cognitive science.

Paper v: The Missing Link Between Memory and Reinforcement Learning

The model presented in this paper builds on the one presented in paper 2, adding a biologically plausible network model for value–based choice. With this extension, we show how memory can support decision–making when value cannot be directly determined solely by immediate observation. The forward–looking aspects of the memory model affords simulation of future scenarios based on a choice as well as the evaluation of those scenarios. The choice–making circuitry supports gradual accumulation of value based on attending to different properties that may be located at different places in the sensory space. The mental time travel aspect of the model thus provides a potential "missing link" between the associative nature of reinforcement learning, and semantic– and episodic memory. In the context of foraging, this paper considers how spatial attention, and prospective simulation based on experience can support complex choice processes.

Result: Simulations of noisy choice processes between two objects, based on a single value–attribute, indicate that increased noise gives faster reaction time, but more random outcomes; reaction time increases as the perceived value of the items become more similar. For simulations involving episodic–like memory, results indicate that goal that is closer in time is reached through fewer episodic transitions, and wins more often. The model also shows how a smaller immediate reward can win out over a larger delayed reward by means of temporal discounting.

Contribution: A computational model of choice processes that shows how episodic memory, attention, and value-based neurological processes can work together to regulate behaviour. The model offers explanation both of simple immediate choices, choices that depend on multiple sensory factors, as well as complicated selections between alternatives requiring forward looking simulations employing episodic and semantic memory structures.

Paper vi: Intelligence as accurate prediction

Intelligence as a concept is often broadly defined to encompass the wide variety of behaviour that purportedly intelligent beings can display, and the complexity of situations with which they can cope. However, broad definitions can make implementation in physical systems problematic. In this paper we argue that intelligence may be usefully approximated as the ability to generate accurate predictions. The argumentation synthesizes Godfrey–Smith's environmental complexity theory, and Hohwy's predictive mind theory, but adding the notion of niche broadness, as well as changes concerning the view of cognition and control. We make explicit the significance of accuracy as a composite of trueness and precision, where the nervous system acts as a distributed controller motivating actions that keep the body in homeostasis. In relation to foraging, accurate prediction thus affords problem solving in general, and obstacle negotiation in particular.

Contribution: A concretization of the intelligence concept that is commensurate with cognitive science and the predictive processing framework.

Paper vii: Direct approach or detour – a comparative model of inhibition and neural ensemble size in behavior selection

This paper presents a biologically plausible model of approach behaviour using spiking units, and inspired by anatomical pathways in mammalian brains that mediate strategy switching between direct approach and approach via detours. The model controls an agent that moves in real–time in a 3D virtual world that is designed to be similar to detour experiments using animals. Detour behaviour for the agent is mediated by the use of a spatial, allocentric map on which alternative paths can be plotted. This corresponds to purported employment of grid–cell populations in animals. We compare the simulated data with data from animal studies, finding that our results are qualitatively similar. In a foraging context, animals are often presented with obstacles that necessitates detour behaviour, and the ability to cope with detours without being distracted tends to be positively associated with cognitive capacity.

Result: Varying inhibition and neural population sizes over simulations indicate that both choice and inhibition processes are facilitated by increasing number of neural units. Larger populations are also associated with reduced behavioural variance and the ability to maintain a detour path without being distracted by a visible goal.

Contribution: Shows a relation between behaviour and neural population size. This work also contributes a biologically realistic computational model of approach behaviour that acts in a virtual 3D world, and which makes up a framework for experimental work.

Papers to which the author contributed, not part of this thesis

Bååth, R., Tjøstheim, T. A., & Lingonblad, M. (2016). The role of executive control in rhythmic timing at different tempi. *Psychonomic bulletin & review*, 23(6), 1954-1960.

Balkenius, C., Tjøstheim, T. A., & Johansson, B. (2019, March). Arousal and Awareness in a Humanoid Robot. In *AAAI Spring Symposium: Towards Conscious AI Systems*.

Tärning, B., & Tjøstheim, T. A. (2019, September). Gender differences in allocation of attention and read time in an educational history game. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (pp. 229-231).

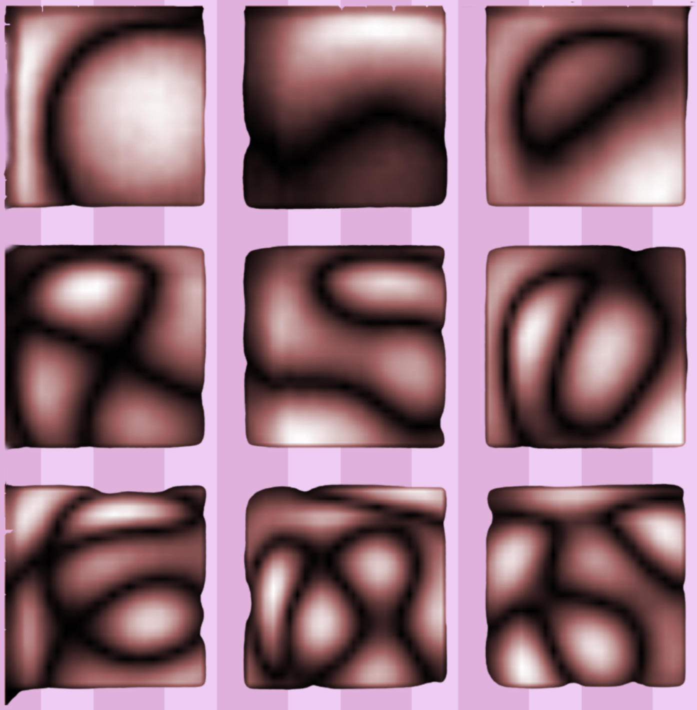
Balkenius, C., Johansson, B., & Tjøstheim, T. A. (2020). Ikaros: a framework for controlling robots with system-level brain models. *International Journal of Advanced Robotic Systems*, 17(3), 1729881420925002.

Johansson, B., Tjøstheim, T. A., & Balkenius, C. (2020). Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, 17(2), 1729881420911498.

Tjøstheim, T. A., Stephens, A., Anikin, A., & Schwaninger, A. (2020). The cognitive philosophy of communication. *Philosophies*, 5(4), 39.

Stephens, A., Tjøstheim, T. A., Roszko, M. K., & Olsson, E. J. (2021). A Dynamical Perspective on the Generality Problem. *Acta Analytica*, 36(3), 409-422.

Olsson, E., Tjøstheim, T. A., Stephens, A, Schwaninger, A., Roszko, M. (submitted). The Cognitive Basis of the Conditional Probability Solution to the Value Problem for Reliabilism.



Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

Faculties of Humanities and Theology
Department of Philosophy
Cognitive Science

Lund University Cognitive Studies 183
ISBN 978-91-89415-23-2
ISSN 1101-8453