

Structuring microscopic dynamics with macroscopic feedback

From social insects to artificial intelligence

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt

der Fakultät Physik

Bereich Mathematik und Naturwissenschaften

der Technischen Universität Dresden

von

Adolfo Alsina López

geboren am 16.12.1993 in Melilla, Spanien

Max-Planck-Institut für Physik komplexer Systeme

Dresden, 2021

Eingereicht am Oktober 11, 2021

Gutachter:

1. Prof. Dr. Benjamin Friedrich
2. Dr. Philip Greulich
3. Prof. Dr. Frank Jülicher

I love deadlines. I love the whooshing noise they make as they go by.
Douglas Adams - The Salmon of Doubt

Abstract

Physical processes rely on the transmission of energy and information across scales. In the last century, theoretical tools have been developed in the field of statistical physics to infer macroscopic properties starting from a microscopic description of the system. However, less attention has been devoted to the remodelling of microscopic degrees of freedom by macroscopic feedback. In recent years, ideas from non-equilibrium physics have been applied to characterise biological and artificial intelligence systems. These systems share in common their structure in discrete scales of organisation that perform specialised functions. To correctly regulate these functions, the accurate transmission of information across scales is crucial. In this thesis we study the role of macroscopic feedback in the remodelling of microscopic degrees of freedom in two paradigmatic examples, one taken from the field of biology, the self-organisation of specialisation and plasticity in a social wasp, and one from artificial intelligence, the remodelling of deep neural networks in a stochastic many-particle system.

In the first part of this thesis we study how the primitively social wasp *Polistes canadensis* simultaneously achieves robust specialization and rapid plasticity. Combining a unique experimental strategy correlating time-resolved measurements across vastly different scales with a theoretical approach, we characterise the re-establishment of the social steady state after queen removal. We show that *Polistes* integrates antagonistic processes on multiple scales to distinguish between extrinsic and intrinsic perturbations and thereby achieve both robust specialisation and rapid plasticity. Furthermore, we show that the long-term stability of the social structure relies on the regulation of transcriptional noise by dynamic DNA methylation.

In the second part of this thesis, we ask whether emergent collective interactions can be used to remodel deep neural networks. To this end, we study a paradigmatic stochastic many-particle model where the dynamics are defined by the reaction rates of single particles, given by the output of distinct deep neural networks. The neural networks are in turn dynamically remodelled using deep reinforcement learning depending on the previous history of the system. In particular, we implement this model as a one dimensional stochastic lattice gas. Our results show the formation of two groups of particles that move in opposite directions, diffusively at early times and ballistically over longer time-scales, with the transition between these regimes corresponding to the time-scale of left/right symmetry breaking at

the level of individual particles. Over a hierarchy of characteristic time-scales these particles develop emergent, increasingly complex interactions characterised by short-range repulsion and long-range attraction. As a result, the system asymptotically converges to a regime characterised by the presence of anti-ferromagnetic particle clusters. To conclude, we characterise the impact of memory effects and demographic disorder on the dynamics.

Together, our results shed light on how non-equilibrium systems can employ macroscopic feedback to regulate the propagation of fluctuations across scales.

Zusammenfassung

Physikalische Prozesse beruhen auf der Übertragung von Energie und Information über Skalen hinweg. Im letzten Jahrhundert wurden im Bereich der statistischen Physik theoretische Instrumente entwickelt, um aus einer mikroskopischen Beschreibung des Systems auf makroskopische Eigenschaften zu schließen. Weniger Aufmerksamkeit wurde jedoch der Umgestaltung mikroskopischer Freiheitsgrade durch makroskopische Rückkopplung gewidmet. In den letzten Jahren wurden Ideen aus der Nichtgleichgewichtsphysik angewandt, um biologische und künstliche Intelligenzsysteme zu charakterisieren. Diesen Systemen ist gemeinsam, dass sie auf diskreten Organisationsebenen strukturiert sind und spezialisierte Funktionen ausführen. Um diese Funktionen korrekt zu regulieren, ist die genaue Übertragung von Informationen über Skalen hinweg entscheidend. In dieser Arbeit untersuchen wir die Rolle der makroskopischen Rückkopplung bei der Umgestaltung mikroskopischer Freiheitsgrade an zwei paradigmatischen Beispielen, einem aus dem Bereich der Biologie, der Selbstorganisation von Spezialisierung und Plastizität in einer sozialen Wespe, und einem aus dem Bereich der künstlichen Intelligenz, der Umgestaltung von tiefen neuronalen Netzen in einem stochastischen Vielteilchensystem.

Im ersten Teil dieser Arbeit untersuchen wir, wie die primitiv soziale Wespe *Polistes canadensis* gleichzeitig eine robuste Spezialisierung und schnelle Plastizität erreicht. Durch die Kombination einer einzigartigen experimentellen Strategie, die zeitaufgelöste Messungen über sehr unterschiedliche Skalen hinweg mit einem theoretischen Ansatz kombiniert, charakterisieren wir die Wiederherstellung des sozialen Gleichgewichts nach dem Entfernen der Königin. Wir zeigen, dass *Polistes* antagonistische Prozesse auf mehreren Ebenen integriert, um zwischen extrinsischen und intrinsischen Störungen zu unterscheiden und dadurch sowohl eine robuste Spezialisierung als auch eine schnelle Plastizität zu erreichen. Darüber hinaus zeigen wir, dass die langfristige Stabilität der Sozialstruktur von der Regulierung des Transkriptionsrauschens durch dynamische DNA-Methylierung abhängt.

Im zweiten Teil dieser Arbeit stellen wir die Frage, ob emergente kollektive Interaktionen genutzt werden können, um tiefe neuronale Netze umzugestalten. Zu diesem Zweck untersuchen wir ein paradigmatisches stochastisches Vielteilchenmodell, bei dem die Dynamik durch

die Reaktionsraten einzelner Teilchen definiert ist, die durch den Output verschiedener tiefer neuronaler Netze gegeben sind. Die neuronalen Netze werden ihrerseits durch tiefes Verstärkungslernen in Abhängigkeit von der Vorgeschichte des Systems dynamisch umgestaltet. Insbesondere implementieren wir dieses Modell als ein eindimensionales stochastisches Gittergas. Unsere Ergebnisse zeigen, dass sich zwei Gruppen von Teilchen bilden, die sich in entgegengesetzte Richtungen bewegen, und zwar diffus zu frühen Zeiten und ballistisch über längere Zeitskalen, wobei der Übergang zwischen diesen beiden Regimen der Zeitskala des Links/Rechts-Symmetriebrechens auf der Ebene der einzelnen Teilchen entspricht. Über eine Hierarchie charakteristischer Zeitskalen entwickeln diese Teilchen emergente, zunehmend komplexe Wechselwirkungen, die durch kurzreichweitige Abstoßung und weitreichende Anziehung gekennzeichnet sind. Infolgedessen konvergiert das System asymptotisch zu einem Zustand, der durch das Vorhandensein von antiferromagnetischen Teilchenclustern gekennzeichnet ist. Abschließend charakterisieren wir die Auswirkungen von Gedächtniseffekten und demografischer Unordnung auf die Dynamik.

Insgesamt werfen unsere Ergebnisse ein Licht darauf, wie Nichtgleichgewichtssysteme makroskopische Rückkopplungen nutzen können, um die Ausbreitung von Fluktuationen über Skalen hinweg zu regulieren.

Acknowledgments

Four years ago I started my PhD without understanding the true importance of the acknowledgments section of a thesis. Back then I thought that they were something that one could omit, as the contents of a thesis are supposed to be the result of the intellectual effort of its author. However, now I came to understand that they are a beautiful glimpse at the life of the author during the four years of their PhD. These lines are a window into my personal and professional life dedicated to all the people who have supported me during this time. There is not enough space in here to thank all the people who deserve it, but I will nevertheless try my best to do so.

First, I want to state the obvious, the Rulands group has been a great environment full of wonderful people. I would not have become the scientist that I am today without their advice. The original members of the group, Bahareh, Fabian and Fabrizio, had to endure my first attempts at giving talks: confusing, long and also certainly boring. Luckily for them, I actually learned something from their comments (what a long way we've come!), making the life of the new members easier. Álvaro, Matteo and Yiteng have all fit seamlessly in the group and became part of many great memories.

Although the last two years have been certainly strange, Felix and Aida made the lockdown easier with our weekly scientific discussions. Our meetings kept me motivated and helped me look at my projects from different points of view, revealing new questions and answers that I had not considered myself. I also want to mention visitors like Najme, with her permanent smile and insatiable curiosity, that helped me expand my scientific horizons during their time at PKS. Finally, the opportunity of working with Misha and Onurcan, two brilliant students, was extremely special to me. The experience of supervising someone taught me a lot about how to properly communicate my ideas and to define clear, attainable research goals.

But any mention to the Rulands group would be incomplete without its founder. Steffen, during these four years you have been a wonderful supervisor who has taught me how to be a scientist. You have led us by example with your constant dedication. You have always been available for us and offered us valuable advice whenever we needed it. I will always remember and cherish my time at the Rulands group during the next steps of my career.

During these four years our group has been part of the Biological Physics division of the MPI-PKS. I would like to thank all the present and past members of the division, as their input during seminars, informal meetings and even lunch, has helped me shape the contents of this thesis. In particular I would like to thank Fabrizio, Felix, Stefano and Yiteng for reviewing parts of this document and providing feedback. The non-scientific staff at the division, particularly Anna and Ulrike, and the IT department have made my life much easier during my time at PKS, allowing me to focus on science as much as possible. To conclude with the scientific side, I could not avoid mentioning Solenn Patalano, a wonderful collaborator that made me discover a fascinating system that I had never expected to work on.

However, becoming part of this environment is just another step in a long process that started more than ten years ago. Now I would like to specially thank three people that have been instrumental in shaping my professional life. Without their support and encouragement it is likely that my career path would have been completely different. Salva, you and your relentless enthusiasm and passion for teaching made me fall in love with this discipline at a time where I did not know what I wanted to do. I would not have believed it if someone told me at the time than ten years later I would be dedicating my thesis to you, yet here I am. Serge, you made me realise how fascinating scientific research can be. Thanks to your drive and patience, you motivated me to pursue a career in fundamental research, a dream that now is closer to becoming a reality. Finally, Martin, the joy you transmitted during your lectures was the force driving me to the university every Monday morning and it ended up infusing me with passion for statistical mechanics. Without all of you this would not have been possible.

Lastly, no acknowledgments section of any thesis would be complete without thanking the arguably most important people of all, my friends and my family. Amanda, Ana, Charlie, Dora, Fabrizio, Inés, Lucía, Omar, Piotr, Roberto, Sara, Stefano, Talía. You all made my life better in one way or another during these last few years. Amanda, Dora, Omar, Piotr, Stefano, Talía, having dinner with you, our nights singing in an improvised karaoke or dancing italodisco and our discussions about politics or Finnish movies made these four years in a cold East German city unforgettable. Ana, the first person I have ever met that had been to Melilla. Since the night we met we have not stopped laughing, an unmistakable sign that our friendship is one for the ages. Hopefully in the near future we can celebrate our defenses together in a nice sunny place. Charlie, I could not have asked for a better office partner. You have been a true role model at both the personal and professional levels, single-handedly making every event better with your presence. Your friendship is one of the most valuable

things that I will keep from my time in Dresden. And, as if this was not enough, you have also introduced me to what has quickly become one of my favourite hobbies. Fabrizio, although we started our PhDs at the same time, I have learned an enormous amount from you, you are a wonderful scientist and it has been great sharing this process with you. The day you got into chess, I was equally stoked and surprised. I hope that you keep improving in the future but that, nonetheless, the student never surpasses the master. Lucía and Sara, old friends by now, you have seen me grow during the last four years, but not everything has been easy for you. I need to thank you for your understanding and always being there for me when I needed to talk to someone. Roberto, our paths have run parallel for the last four years. Since we were all living in the guest house to now, we have lived many great moments together. Being in the same situation has also made this writing months feel much easier than they would have otherwise been.

Inés, you know why you are the last one, *lo mejor viene al final*. You have changed my life completely. You made me realise that nothing makes sense without the proper work-life balance. You have supported me immensely and have always brought me back in touch with reality in those days when I feel that nothing works. This is just the beginning of a wonderful story and I cannot wait to see what the future has in store for both of us. To us.

Finally, my family. Without their support I would have probably never left the small piece of land where I was born. Since I was a child they have constantly encouraged me to learn for the sake of learning, giving me at the same time absolute freedom to decide what I wanted to do with my life. *Sin vosotros nada de esto sería posible. Siempre habéis apoyado mis constantes mudanzas, mis decisiones arriesgadas y mi interés por explorar nuevos lugares aunque a veces no os hiciera mucha gracia. Poder acabar esta tesis es el final de un proceso que empezasteis vosotros cuando yo todavía era un niño. Por todo esto y mucho más, gracias.*

Contents

1. Introduction	1
1.1. Biological systems are organised in a discrete hierarchy of spatial scales	2
1.2. The regulation of biological function after perturbations relies on phenotypic plasticity	3
1.2.1. Phenotypic plasticity in social insects	4
1.2.2. Theoretical approaches to phenotypic plasticity	5
1.3. Robustness of artificial intelligence systems against noise	6
1.3.1. Deep neural networks can be dynamically remodelled after a perturbation	8
1.4. Outline of the thesis	10
2. Emergence of specialisation and plasticity in a social insect	11
2.1. Experimental induction of the reprogramming process	13
2.2. Characterization of molecular-level processes using genomics	15
2.3. Symmetry breaking of molecular states	18
2.4. Master equation description of the multiscale dynamics	23
2.5. Phase diagram of the system	24
2.6. Mean-field master equation description of the nest dynamics	27
2.7. Continuum limit of the mean-field master equation and phase portrait	30
2.8. Quantitative prediction of experimental observables	36
2.8.1. Marginal distribution of ovary sizes	36
2.8.2. Quantification of the activity stemming from fighting interactions . . .	40
2.8.3. Estimation of interaction parameters and associated timescales	41
2.9. Epigenetic regulation leads to an increase of the lifetime of the society by reducing gene expression variance	41
2.10. Conclusions	48
3. Driving stochastic lattice systems by reinforcement learning	51
3.1. An introduction to deep reinforcement learning	55
3.1.1. Reinforcement learning	55
3.1.2. Deep learning and its application to reinforcement learning tasks . . .	58

3.2. Deep reinforcement learning driven lattice gas	62
3.2.1. Details of the numerical implementation	64
3.3. Collective dynamics emerge as a result of interactions induced by the shared environment	66
3.4. Quantification of the dynamics at the macroscopic level	71
3.5. Characterization of the neural network dynamics	73
3.6. Emergence of complex interactions as a result of the interplay between the macroscopic and the microscopic scales	78
3.7. Increasing the discount parameter induces a transition between asymptotic dynamical regimes	79
3.7.1. Autocorrelation functions of the individual velocities	82
3.8. Effect of demographic disorder on the collective dynamics	84
3.9. Conclusions	85
4. Conclusions and future perspectives	91
Bibliography	97
A. Numerical simulations of the master equations	109

1. Introduction

Denn kein Gedicht gilt dem Leser, kein Bild dem Beschauer, keine Symphonie der Hörschaft.

Walter Benjamin - Die Aufgabe des Übersetzers (Illuminationen)

Physical processes rely on the transmission of energy and information across scales. A striking example is that of turbulent flows, where energy is transferred across a wide range of spatial scales, a phenomenon known as energy cascade [1–3]. One of the longstanding goals of physics has been to understand how macroscopic structures emerge as a result of interactions between their fundamental microscopic components.

The field of statistical physics has witnessed during the last century the development of powerful theories designed to provide answers to this question. For instance, one of the cornerstones of modern physics, the Renormalization group (RG) [4–6], leverages the power of scale transformations to obtain information about the large scale behaviour of systems close to a critical point. In particular, fixed points of the RG flows define universality classes, collections of models that, although they differ in their microscopic details, exhibit the same properties in a scaling limit [7–9].

However, whereas research has traditionally focused on the emergence of macroscopic structures from microscopic constituents, the effect of macroscopic feedback on the remodelling of microscopic degrees of freedom has received less attention. Furthermore, in recent years research on non-equilibrium systems, in particular in the field of active matter [10–12], has increased, fueled by experimental advances. Although these systems share in common their constant consumption of energy at the microscopic scale, they do not have an unified theoretical description, hindering the study of their microscopic and macroscopic properties.

The goal of this thesis is to study the role of macroscopic feedback in the remodelling of microscopic degrees of freedom in non-equilibrium systems, drawing lessons from the fields of biology, in particular from social insects, and artificial intelligence. In this chapter we introduce the biological and artificial systems that we consider in Chapter 2 and Chapter 3 and motivate why they are well suited to providing answers to this question.

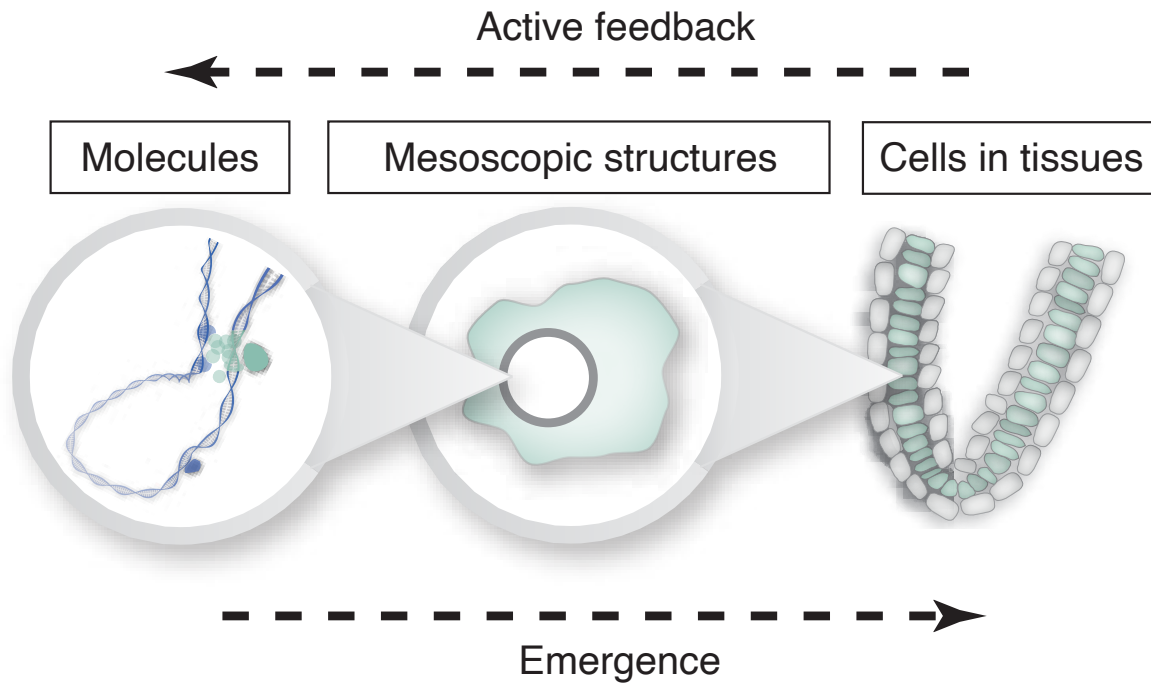


Figure 1.1.: Cartoon illustrating the interplay between different levels of biological organisation.

1.1. Biological systems are organised in a discrete hierarchy of spatial scales

Biological systems are paradigmatic examples of non-equilibrium active matter. They operate far from thermal equilibrium, relying on a continuous influx of energy to form and maintain complex structures that perform specialized functions. As Erwin Schrödinger aptly put it in his seminal lecture “What is life” [13] more than one hundred years ago:

A living organism continually increases its entropy – or, as you may say, produces positive entropy – and thus tends to approach the dangerous state of maximum entropy, which is death. It can only keep aloof from it, i.e. alive, by continually drawing from its environment negative entropy.

In order for biological systems to operate in such a noisy environment their dynamics have to be tightly regulated, as the correct regulation of function is crucial for the survival of the organism. The paramount importance of function in biology is reflected in one of the main features of biological systems, the existence of discrete levels of organisation at different spatial scales, associated to structures such as molecules, cells or tissues, that perform specific functions. It is precisely the connection between structure and function what makes the study of these structures a central question in biological research.

Therefore, in order to operate correctly, biological systems rely on accurate transmission

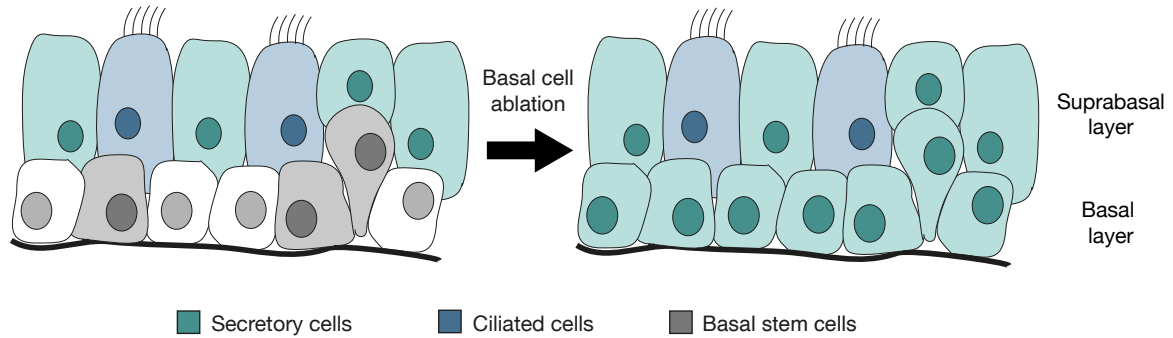


Figure 1.2.: Reprogramming of differentiated cells in the lung epithelium after ablation of the basal stem cell layer. Adapted by permission from Springer Nature: Nature (Dedifferentiation of committed epithelial cells into stem cells *in vivo*, Tata, P.R. *et al.*), (2013) [15]

of information across different levels of biological organisation. As a result, the dynamics at a single scale are regulated by intricate interactions across multiple spatial scales. For instance, Figure 1.1 schematically represents the complex interplay between three levels of this hierarchical structure: molecules that organise into cells, that in turn are structured into tissues.

Although traditionally the focus of biophysical research has been on how biological systems form and maintain complex structures in noisy environments, in recent years, the remodelling of such structures after a perturbation has sparked experimental (Figure 1.2) and theoretical interest [14–16]. However, robust specialization in a noisy environment and rapid plasticity against perturbations cannot be simultaneously achieved relying on bistability alone, as in that case the stability of the system against fluctuations mitigates the rapid reestablishment of lost states [17–19].

In this thesis we study how this dilemma could be resolved using feedback across different scales of spatial organisation. In particular, in the first part of this thesis we will study how a social insect, a primitively social paper wasp, exploits interactions across multiple spatial scales to simultaneously achieve robust specialisation and rapid phenotypic plasticity. In the next section we will introduce the concept of phenotypic plasticity, laying the ground for the work that will be presented in Chapter 2.

1.2. The regulation of biological function after perturbations relies on phenotypic plasticity

Phenotypic plasticity is the ability of organisms to exhibit more than one phenotype from a single genotype in response to environmental changes [20, 21]. A fundamental property

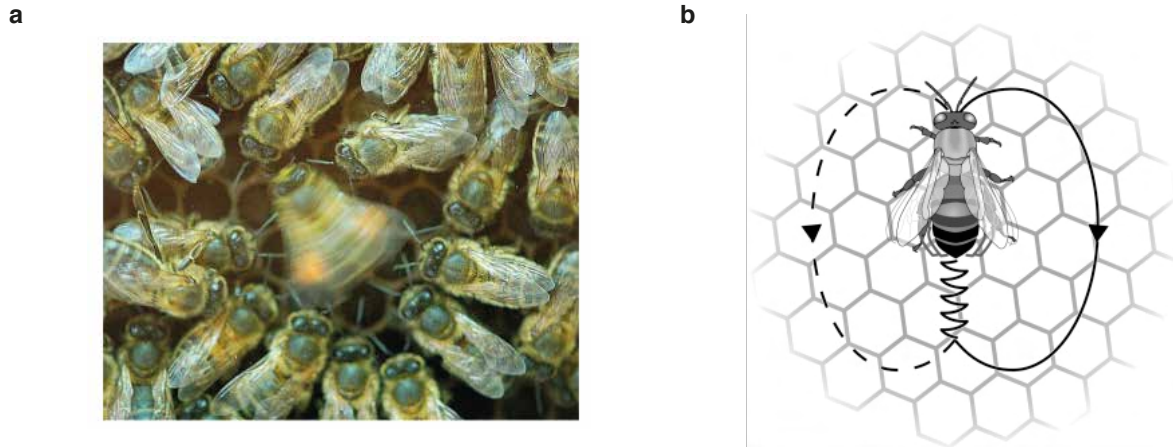


Figure 1.3.: **a)** Honeybee performing a waggle dance in front of a crowd of observer bees. Adapted from [28]. **b)** Schematic description of honeybee movements during a waggle dance. Adapted from [29].

for the survival of organisms under environmental stress, examples of phenotypic plasticity can be found in both the animal and the plant kingdoms [22, 23]. Throughout this thesis we differentiate phenotypic plasticity from developmental plasticity [24, 25], understood as the ability of an organism to produce different phenotypes from a single genotype during development depending on external signals.

In particular, phenotypic plasticity encompasses the plastic response of social organisms against changes in social interactions [23, 26], thus playing a crucial role in the regulation of the social structure. A fascinating example of phenotypic plasticity in social groups is the regulation of caste, corresponding for instance to reproductive and non-reproductive phenotypes, in colonies of social insects. Specifically, alterations in the structure of the society can induce caste changes at the level of individual insects.

In the next section we provide a general introduction to phenotypic plasticity in social insects. We refer the reader to the corresponding chapter for details about *Polistes canadensis*, the primitively social wasp that we study in Chapter 2.

1.2.1. Phenotypic plasticity in social insects

Scientists have been fascinated by the behaviour of social insects for centuries. In the early twentieth century, Karl Von Frisch decoded the complex dances (Figure 1.3) used by foraging honeybees to communicate accurate positional information to other members of the colony [27]. As a result of this discovery, Von Frisch would be awarded the Nobel Prize of Physiology in 1973 together with Nikolaas Tinbergen and Konrad Lorenz for their studies of social behavioural patterns in animals.

The bee dance exemplifies the central role played by social interactions in the regulation

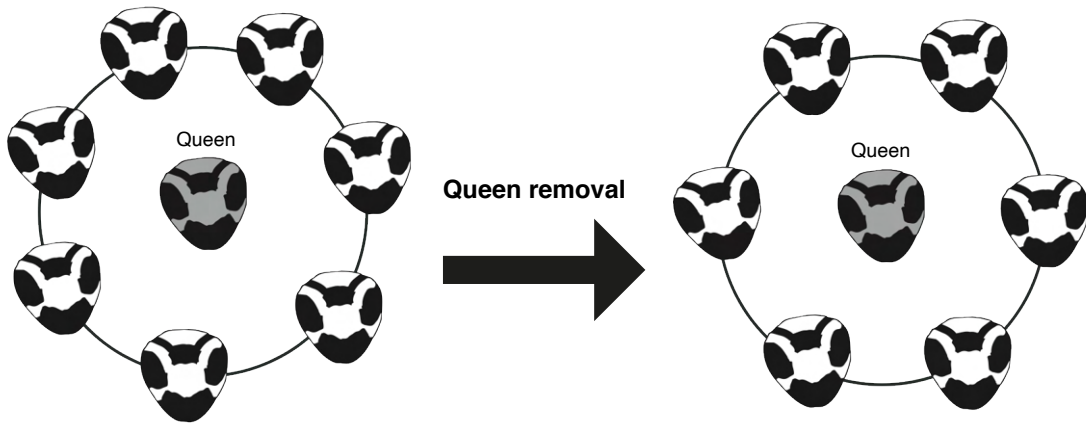


Figure 1.4.: Cartoon illustrating the queen replacement after queen removal in *Polistes canadensis* wasps.

of the colony dynamics. In several cases these interactions lead to the emergence of collective properties, ranging from mechanical properties of insect aggregates [30, 31] to collective dynamics at the behavioural level [32, 33].

The definition of social insects encompasses very different species, as reflected by the number of individuals that form part of a colony, spanning several orders of magnitude, ranging from societies composed of a few individuals to others formed of tens of thousands. In particular, species of social insects are classified according to different sociality levels, ranging from solitary to eusocial (fully social), reflecting the aspects in which they differ from fully social groups. Particularly interesting for us are primitively social insects, a class that only differs from eusocial insects in the fact that primitively social insects exhibit little or no morphological differences between castes. In some species of primitively social insects phenotypic plasticity manifests itself as the ability of single insects to change castes after receiving the right environmental signal [34–36]. In particular, *Polistes canadensis*, a primitively social paper wasp, exhibits caste changes between a non-reproductive and a reproductive phenotype [35, 37]. After queen removal, non-reproductive *Polistes* workers can reprogram, altering their gene expression profile and developing mature eggs, leading to the emergence of a new queen (Figure 1.4). In the first part of this thesis we study how specialization and plasticity in *Polistes* emerge from the interplay between macroscopic and microscopic processes.

1.2.2. Theoretical approaches to phenotypic plasticity

Theoretical and computational studies of phenotypic plasticity have mainly focused on the steady-state or asymptotic properties of the system [38–40]. In these studies, the probability of displaying a given phenotype depends on the current environmental condition that acts as

a control parameter.

This approach presents two main shortcomings to describe the regulation of phenotypic plasticity in social insects. First, although in these models the phenotype is a function of the environmental condition, the environment is independent of the distribution of phenotypes across the population. However, the environment that social insects are exposed to encompasses social interactions, that are modified when the social structure is perturbed. Therefore, a theory of phenotypic plasticity in social insects must take into account the self-organised nature of the environment. Secondly, in order for multiple phenotypes to be stable for the same environmental signal, the fraction of these phenotypes in the population has to be controlled. For instance, established colonies of *Polistes canadensis* are composed of a single queen and several workers, indicating that both phenotypes are stable in the steady state. However, after queen removal a new queen emerges as a result of worker reprogramming, meaning that the worker phenotype is no longer stable after queen removal.

Therefore, in order to understand the regulation of specialization and plasticity in *Polistes canadensis* an approach that takes into consideration these two properties is needed. In Chapter 2 we develop an approach that relies on a minimal description of the microscopic dynamics of the system that reveals how the regulation of robust specialization and rapid plasticity in *Polistes canadensis* results from antagonistic processes at the molecular and the population scales.

The response of these insects to different perturbations relies on the interplay between processes at different spatial scales. Controlling fluctuations and characterising the response of the system against perturbations are central research topics in several fields, not only in biology. For instance, in the field of artificial intelligence it is desired that systems are robust against fluctuations in the input data and against changes in their internal structure. Inspired by our findings, in the second part of this thesis we studied the role of macroscopic feedback in the remodelling of artificial intelligence systems. In the remainder of this introduction we define artificial intelligence and discuss how artificial neural networks can be used to study the remodelling of microscopic degrees of freedom by macroscopic feedback in an artificial intelligence system.

1.3. Robustness of artificial intelligence systems against noise

Artificial intelligence (AI) is a field of research devoted to the study of intelligent agents, artificial systems that mimic human intelligence to maximize their probability of achieving a certain goal. Nowadays, AI is a thriving field of research in theoretical and applied computer

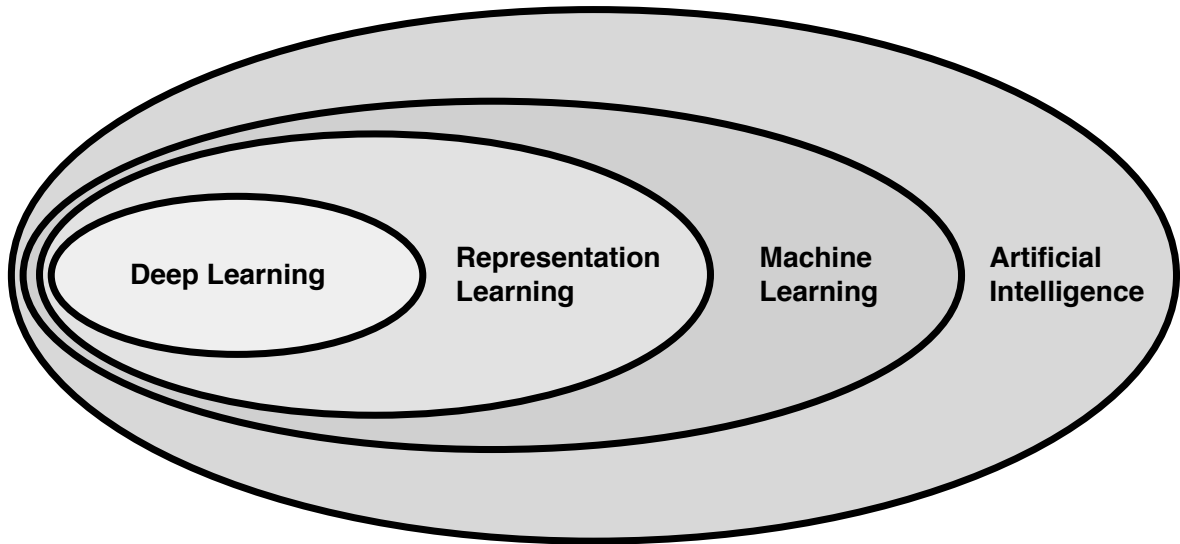


Figure 1.5.: Venn diagram of the different subfields of artificial intelligence.

science, driven by the success of the theory of statistical machine learning and deep artificial neural networks [41–43].

Since its inception as a field, AI methods have been successfully applied to solve problems that could be defined using a list of rules, as demonstrated by IBM Deep-Blue defeating world chess champion Garry Kasparov in 1997. However, AI algorithms did not perform well on problems that are not relying on such a list of rules, such as image recognition.

A subfield of artificial intelligence (Figure 1.5), machine learning refers to methods that infer probability distributions from data. In particular, we refer as representation learning to a subset of machine learning techniques concerning the automatic extraction of features from data, without requiring manual input. Not relying on hard-coded rules (Figure 1.6), machine and representation learning algorithms have enjoyed success in problems where traditional AI algorithms did not perform well, such as classification tasks [44–46].

In the last twenty years, machine learning algorithms have been applied to complex image and voice recognition tasks, recommendation tasks and even autonomous driving, powered by artificial deep neural networks. The field that studies algorithms based on artificial deep neural networks is known as deep learning. However, despite its successes, the predictions of deep learning models can be very sensitive to small perturbations of the input data [47], also known as adversarial examples. Furthermore, machine learning models are affected by catastrophic forgetting [48], a name that refers to the loss of accuracy on old patterns when the system is trained on a new set.

In analogy to biological systems, structure (the network architecture) and function (the response to a given set of input patterns) are closely related in artificial neural networks. In

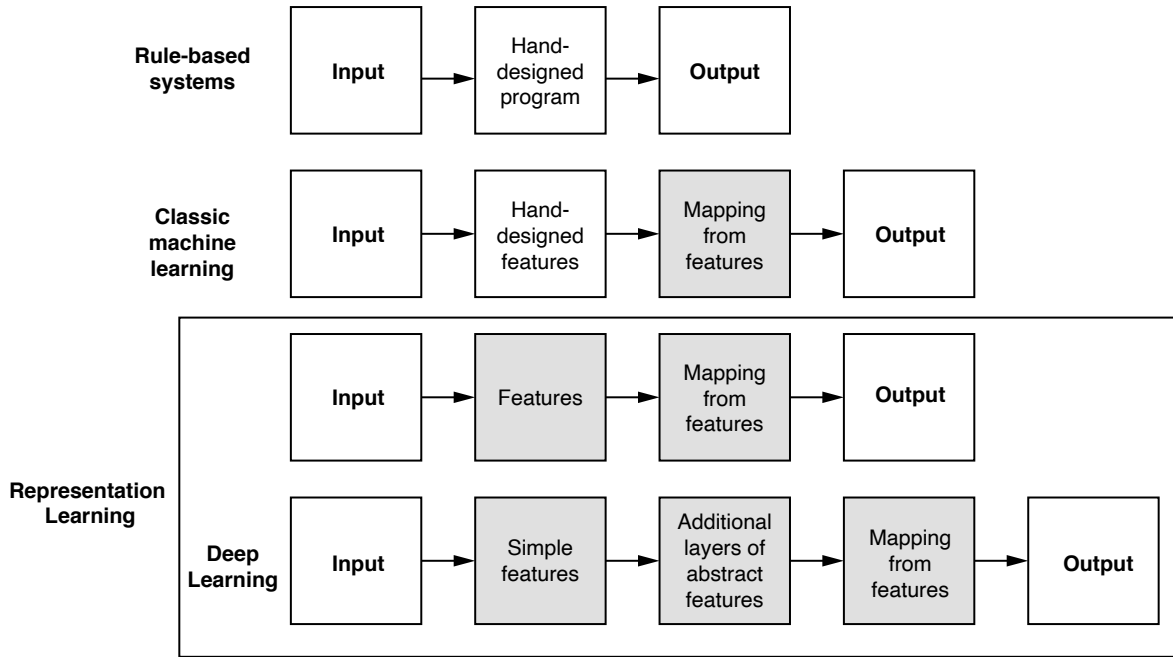


Figure 1.6.: Flowcharts indicating the relationships between different system components within different AI subfields. Shaded boxes indicate components that extract information from data. Adapted from [41].

order for the system to operate correctly, the connection between the two has to be tightly regulated. In the second part of this thesis we explore the relationship between structure and function in a system powered by artificial neural networks that is driven by macroscopic feedback using reinforcement learning.

In the next section we introduce artificial neural networks and reinforcement learning as a learning paradigm. As in Chapter 3 we provide an in-depth introduction to the deep learning concepts and techniques relevant for this thesis, in this introduction we limit ourselves to introducing artificial deep neural networks.

1.3.1. Deep neural networks can be dynamically remodelled after a perturbation

Inspired by the structure of the human brain, artificial neural networks are computing systems composed of connected units. Each individual unit receives information from the rest of the network, producing a signal as a result. The output of a unit is a non-linear function of the weighted sum of its inputs. In particular, the output of unit i can be written as

$$y_i = f \left(\sum_j w_{ij} x_j + b_i \right), \quad (1.1)$$

where y_i is the output of unit i , x_j is the input to unit i from unit j , w_{ij} are the weights connecting units i and j , b_i is a constant representing the bias of unit i and f is a non-linear

function, known as the activation function.

The power of artificial neural networks resides in their ability to be trained from data to infer probability distributions. Specifically, training a neural network corresponds to controlling its output by modifying its set of parameters $\theta = (\mathbf{w}, \vec{b})$, i.e. weights and biases. To successfully train neural networks, their performance is evaluated by means of a loss function, as for instance the cross-entropy [49, 50] between the empirical and the inferred distributions, that is minimized during training.

Depending on how neural networks are trained, deep learning algorithms are divided in three main classes: unsupervised learning, supervised learning and semi-supervised learning, the class to which reinforcement learning belongs [51]. First, in unsupervised learning networks are trained using datasets that do not contain explicit information about any of the data points. For this reason, unsupervised learning algorithms are typically used to extract features from data [41–43]. On the other hand, supervised learning algorithms provide the network with labelled data during the training process. As a result, a trained network learns how to classify unlabelled samples taken from the same probability distribution [41–43].

In a reinforcement learning problem, networks learn from the past history of the system. In particular, networks receive a reward based on their output and are trained to maximize the cumulative reward. In many applications of reinforcement learning [52, 53], the network represents an agent that interacts with an environment and the output of the network corresponds to the actions that the agent can take. As a result of these actions, the environment is modified and the agent receives a reward. In this situation, the network dynamically generates its own training dataset from experience, not requiring a curated set of examples for training.

Therefore, artificial neural networks trained using reinforcement learning provide a suitable setting to explore the interplay between microscopic and macroscopic degrees of freedom in artificial intelligence. In particular, microscopic degrees of freedom correspond to the parametric degrees of freedom of neural networks. These are in turn remodelled during training using reinforcement learning and a dataset that is self-generated by the networks themselves. As the training dataset is created from the output of the neural networks, an effective feedback loop is induced between the microscopic and the macroscopic scales.

In particular, in the second part of this thesis we study how macroscopic feedback remodels microscopic degrees of freedom in a system composed of multiple neural networks trained by reinforcement learning. The neural networks share the same environment, that is in turn modified by their actions. As in reinforcement learning the environment determines the input

of the neural networks, the dynamics of single networks depend on the dynamics of the rest of the system. Therefore, such a system permits us to uncover the response of microscopic degrees of freedom to macroscopic feedback.

In the next section we outline the structure of this thesis, detailing the contents of the chapters that form part of this dissertation.

1.4. Outline of the thesis

In this thesis we study the remodelling of microscopic degrees of freedom by macroscopic feedback in two non-equilibrium systems inspired by biology and artificial intelligence respectively. The thesis is divided in two parts corresponding to these two fields of research.

In the first part of this thesis (Chapter 2) we characterise the interplay between macroscopic and microscopic degrees of freedom in the regulation of an insect colony. In particular, we study how biological systems can simultaneously achieve robust specialization in noisy environments whilst at the same time exhibiting rapid phenotypic plasticity after perturbations. Using the primitively social paper wasp *Polistes canadensis* as a model system, we combine a multiscale experimental approach with a theoretical description using methods from the theory of stochastic processes and dynamical systems.

In the second part of this thesis (Chapter 3) we study, making use of deep learning tools, a many-particle system with a large number of intrinsic variables to explicitly explore the remodelling of microscopic degrees of freedom by macroscopic feedback. In particular, we consider a stochastic many-body system where the dynamics are defined by the reaction rates of individual particles that are in turn determined by deep neural networks trained using reinforcement learning. The characterization of the macroscopic and microscopic dynamics reveals the interplay between the two scales in this minimal model.

Lastly, we conclude in Chapter 4 with some general remarks and a brief discussion of future research perspectives.

2. Emergence of specialisation and plasticity in a social insect

If I be waspish, best beware my sting.

William Shakespeare - The Taming of the Shrew

Biological systems have the remarkable capacity of building complex spatio-temporal structures that are robust over long periods of time in noisy environments. The formation of these structures is governed by processes taking place on vastly different spatial scales, ranging from the molecular to the tissue or population-levels. Traditionally, the focus of theoretical and experimental biophysical research has been centered on understanding the underlying processes regulating the formation of complex structures [31, 54].

However, in recent years it has become clear that biological systems also have the remarkable capacity to break up and rebuild these structures [15, 55, 56], efficiently restoring the steady state of the system after a large perturbation. The regulation of specialisation and plasticity in primitively social insects is an example of such behaviour.

In particular, colonies of social insects rely on the long-term specialisation of individuals into distinct castes, such as queens and workers [24]. Although such phenotypes can be stable over long periods of time, limited by the lifetime of these insects, in the face of environmental noise individuals are nevertheless capable of being rapidly phenotypically reprogrammed:



Figure 2.1.: Nest of *Polistes canadensis*. Picture from Solenn Patalano.

upon receipt of specific cues they undergo a transient phase in which an individual's behaviour normally associated with a specific caste is rapidly altered in order to perform a different task from the one it performed initially (plasticity) [57–59].

In physical terms, the regulation of specialisation typically relies on the establishment of distinct stable states (bistability) which is often mediated by positive feedback loops [18]. In bistable systems, transitions between stable phenotypic states require the crossing of a potential or entropic barrier with a typical crossing time that depends exponentially in the ratio between the energy difference between the two states ΔV and the noise strength η , $T \approx e^{-\Delta V/\eta}$ [17, 19]. Therefore, the same barrier that stabilises the system, mitigates against the rapid reestablishment of a lost phenotype.

Nevertheless, this reasoning appears to contradict the rapid reprogramming observed in plastic systems. Given that the specialised states are stable over long times, this sort of mechanism would lead to reprogramming times much longer than empirically observed. How is it then possible to simultaneously achieve robust specialisation and rapid plasticity against specific environmental changes? As bistability alone cannot provide the answer to this question, we hypothesised that in order to simultaneously achieve these two seemingly antagonistic goals, feedback across different scales of biological organisation might give rise to additional effective degrees of freedom which allow for plasticity while maintaining stable specialisation in the steady state.

To test this hypothesis, we used a well-established model system of phenotypic plasticity, the primitively social paper wasp *Polistes canadensis* (Figure 2.1). After the emergence of the queen's daughters (workers), a stable colony of paper wasps is established, with a single reproductive queen and 8-30 non-reproductive workers (specialisation) [35, 37, 60–63]. These insects display strong phenotypic plasticity, as reflected by the rapid reprogramming of the remaining workers after queen death (or experimental queen removal). In particular, reprogramming was experimentally induced by our experimental collaborators by removing the queen from stable nests and the relaxation dynamics of the system were tracked across multiple scales until the restoration of the steady state. One of the main advantages of choosing *Polistes* as our model system is that it provides the unique opportunity of simultaneously tracking the relaxation dynamics after such a perturbation back to the steady state across multiple spatial scales, therefore allowing us to correlate measurements from the population and the molecular scales at the level of single individuals. Based on these experiments, we then set out to define a model that describes the population and molecular level processes that regulate the reprogramming dynamics and that allows us to understand specialisation

and plasticity as an emergent property thereof. The details of the experimental strategy can be found in Section 2.1.

The main synthesis of our experimental and theoretical approaches shows that *Polistes* integrate processes on different layers of biological organisation to distinguish between intrinsic perturbations of molecular states, against which they are stable, and extrinsic cues like queen removal affecting the society as a whole, against which they react plastically, therefore simultaneously achieving rapid plasticity and robust specialisation. Specifically, the society undergoes a saddle-node bifurcation governed by the population structure as a bifurcation parameter, thereby simultaneously achieving bistability in the steady state and transient, rapid convergence to the queen phenotype after colony-level perturbations.

Additionally, our theoretical model also predicts that such a society regulated by stochastic interactions would not be stable over periods of time longer than the ratio between the characteristic time scales of the population and molecular processes due to strong fluctuations. To resolve this contradiction we proposed, supported by the experimental evidence, a hypothetical mechanism operating at the molecular scale that leads to an exponential increase in the queen turnover time following a reduction in gene expression variability, opening the possibility of a novel role played by epigenetic modifications at the population scale.

This chapter is structured as follows: first, in Section 2.1 we introduce the multi-scale experimental strategy. Then, in Section 2.2 and Section 2.3 we present the experimental results that motivate our model and introduce its main elements. In Section 2.4 the model is written in the form of a non-Markovian master equation that we then simulate in Section 2.5 to explore the phase space of the system. In order to understand the physical mechanism underlying specialisation and plasticity in this system, in Section 2.6 and Section 2.7 we derive a mean-field master equation from the full non-Markovian dynamics and use it to represent a simplified phase portrait of the system that captures the correct structure of the phase space. Afterwards, in Section 2.8 we compare model predictions and the experimental measurements of several different observables. Finally, in Section 2.9 we study the stability of the social order and propose a possible mechanism to stabilise such a society over long periods of time. A summary of the results presented in this chapter can be found in [64].

2.1. Experimental induction of the reprogramming process

Experimental fieldwork was carried out by Solenn Patalano and her team during 3 expeditions to the area of Punta Galeta, Colon in Panama between 2009 and 2012. We refer the interested reader to the Methods section of [64] for more details about location and collection of the

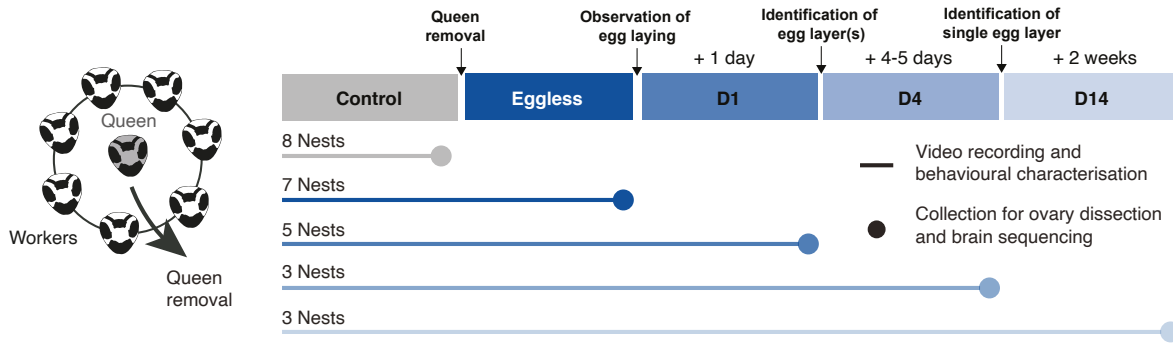


Figure 2.2.: Schematic depiction of the experimental strategy followed in the queen removal experiments.

fieldwork samples.

The experimental approach that was followed is depicted in Figure 2.2. A total of $N = 26$ independent *Polistes* nests were monitored for a period of time of between three and nine weeks until collection. Out of these 26 nests, 8 were collected unmanipulated (Control phase). We empirically define five different phases, corresponding to the different stages of the reprogramming process at which we collect all individuals for further treatment.

- **Control** (8 nests): Nests belonging to this phase are characterised by the presence of a single, well-established queen and several workers.
- **Eggless** (7 nests): The Eggless phase is initiated by the simultaneous removal of the queen and a single egg from a Control nest. All nests collected in this phase were collected before a new egg layer had emerged.
- **D1** (5 nests): Immediately after a new egg appears the Eggless phase was considered to be over. During the following day, the identity of the new egg-layer was determined based on behavioural observations of dominance behaviour.
- **D4** (3 nests): Additional nests were collected between 4 and 5 days after the first egg laying after queen removal had been detected.
- **D14** (3 nests): A final 3 nests were collected two weeks after the first egg laying following queen removal.

Insects were directly collected from the nests using forceps (Figure 2.3a), their heads cut off and immediately placed in solution. Their bodies were conserved in an ethanol solution in freezing temperatures before further analysis. In order to overcome by the destructive nature of the molecular and anatomical measurements, this experimental strategy relies on implicitly assuming the equality of the ensemble and time-averaged statistics.

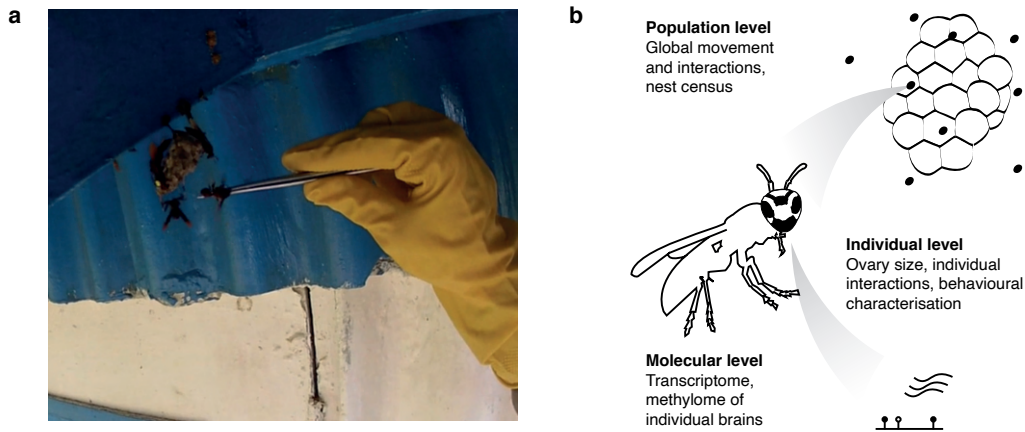


Figure 2.3.: a) Insect removal from a nest using forceps. Picture courtesy of Solenn Patalano. b) Schematic of the multi-scale experimental approach.

This approach is unique because it permits us to track the relaxation dynamics of the system back to the social steady state simultaneously across multiple spatial scales, as illustrated by Figure 2.3. In particular, these measurements inform about three different spatial scales: the population scale, the individual scale and the molecular scale. First, at the population scale the measurements comprise video recordings of several nests, used for behavioural characterization, and nest occupancy census. At the individual level anatomical measurements such as the age or ovary size of individual insects were obtained. Lastly, at the molecular level the transcriptome and methylome of the brain of single insects was sequenced by bulk RNA sequencing and bisulfite sequencing respectively.

2.2. Characterization of molecular-level processes using genomics

One of the main layers of biological regulation, gene expression plays a primary role in the regulation of behaviour and caste of *Polistes* wasps [35, 37, 65]. To quantify differences between queens and workers at the microscopic level we first identified molecular signatures distinguishing reproductive from non-reproductive individuals. Drawing from previous research on *Polistes canadensis* [37], we define as “queen genes” the subset of all differentially expressed genes between reproductive and non-reproductive individuals in Control nests that are expressed by reproductive individuals and not by non-reproductive ones (Figure 2.4a). To elucidate the function of the queen genes, we conducted a gene ontology (GO) analysis of the set of queen genes revealing that these are fundamental for reproductive purposes, as reflected by enrichment of genes that correlate with ovary growth, metabolism and lipid transport, such as *Vitellogenin* and *Apolipoprotein-3* [36, 66, 67].

To clarify the effect of queen removal and replacement on gene expression in workers

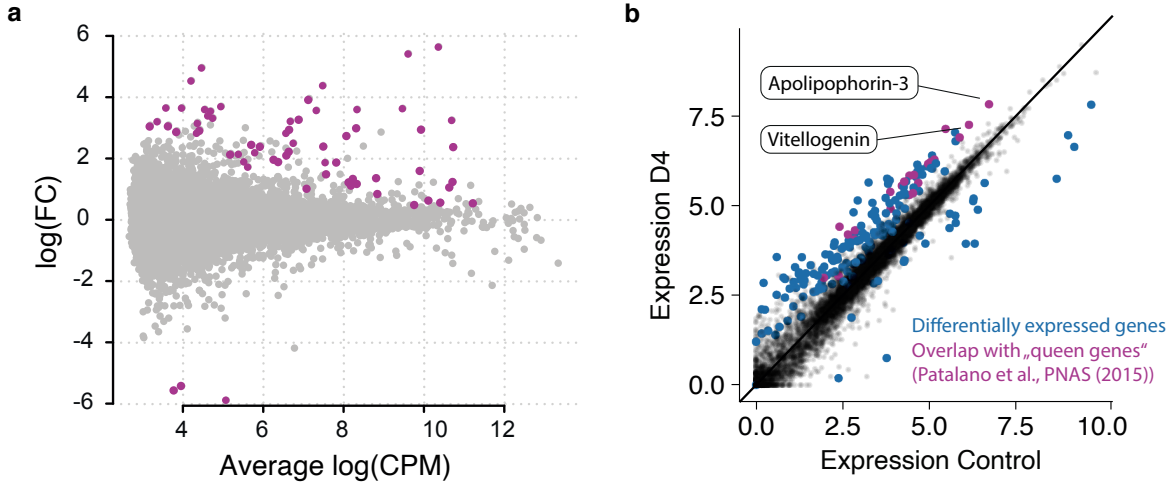


Figure 2.4.: **a)** Log of the ratio between gene expression of reproductive and non-reproductive individuals. Points correspond to individual genes. Color indicates differentially expressed genes between the two groups. Adapted from [37] **b)** Normalized gene expression (RPKM) in Control versus D4 nests. Points correspond to individual genes. Color indicates differentially expressed genes between Control and D4 (blue) and the overlap with the previously defined set of queen genes (purple).

we compared the transcriptomes of workers from Control and D4 nests, finding a total of 227 differentially expressed genes (Figure 2.4b) between the two phases. Queen genes are overrepresented in the set of differentially expressed genes between Control and D4 nests ($p = 3.1 \cdot 10^{-15}$, hypergeometric test), indicating their upregulation following queen removal.

Nevertheless, the dimensionality of the set of queen genes is too large to be amenable to theoretical modelling. To identify possible collective components in the regulation of the queen genes after queen removal we performed dimensionality reduction (Principal Component Analysis - PCA) of the expression of queen genes across different individuals in Control and D4 nests. The PCA analysis (Figure 2.5a) displays a shift in the transcriptomic profiles of D4 insects towards those of established Control queens. Furthermore, a radar plot of the normalized expression of the queen genes (Figure 2.5a) indicates that the upregulation of the queen genes after queen removal is captured by a single principal component, supporting a theoretical description in terms of a single effective degree of freedom representing the collective expression of this set of genes. In the following sections we set out to derive a minimal theoretical description of the reprogramming dynamics, not assuming any non-linear effects unless explicitly motivated by the data.

Finally, in order to quantify to which degree the transcriptome of D4 reproductive individuals resembles that of control queens, we calculated the correlation coefficient $c_{\text{queen},w}$ between the transcriptional profiles of the control queens and single insects, defined as the average between the correlation of the individual expression profiles with those of each of the control

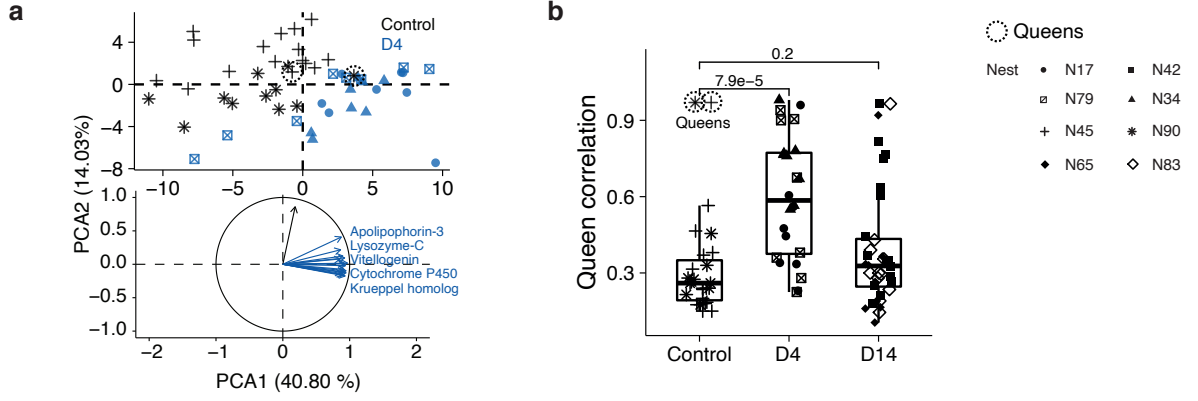


Figure 2.5.: **a)** Principal component analysis of gene expression (top). Dots correspond to single insects. Radar plot of the normalized expression of a selected subset of queen genes (bottom) **b)** Average correlation coefficients of the queen gene expression of individual wasps compared (p-value, t-test) to the two control queens.

queens $c_{queen,w} = (c_{q_1,w} + c_{q_2,w})/2$ (Figure 2.5b), where $c_{q_i,w}$ denotes the Pearson correlation coefficient between the transcriptome of a control queen Q_i with $i \in \{1, 2\}$ and a given insect W ,

$$c_{q_i,w} = \frac{\text{cov}(Q_i, W)}{\sigma_{Q_i} \sigma_W}. \quad (2.1)$$

Together, our findings provide evidence of a collective change in the expression of queen genes of the remaining workers after queen removal, leading to the upregulation of queen genes and an increase of the similarity of their transcriptomic profiles to those of Control queens. Nevertheless, the loss of worker phenotypes is only transient, as reflected by the transcriptomic profiles of insects collected at D14, similar ($p=0.2$, t-test) to those of Control insects.

In the light of the experimental evidence, we set out to derive a biophysical description of the nest dynamics where each insect i will be represented by a single scalar variable $n_i \in \mathbb{N}$, representing the abundance of queen gene products. The evolution of the joint probability of finding queen gene product abundances $\{n_1, \dots, n_{N+1}\}$ across a population of $N + 1$ insects at time t , denoted as $P(\{n_i\}, t)$, depends on processes acting on both the molecular and population scales. Considering for the present only processes taking place at the molecular scale, the simplest model describing the temporal evolution of the nest composition is given by the following master equation

$$\begin{aligned} \frac{d}{dt} P(\{n_k\}, t) = & \sum_{i=1}^{N+1} \mu [P(\{n_i - 1\}, t) - P(\{n_i\}, t)] \\ & + \delta [(n_i + 1)P(\{n_i + 1\}, t) - n_i P(\{n_i\}, t)], \end{aligned} \quad (2.2)$$

where the first two terms represent the production of new queen gene products with constant rate μ and the remaining two terms their poissonian degradation with rate δ .

Within this description, the number of queen gene products in individual insects are statistically independent, leading to a steady state solution of Equation 2.2 given by an homogeneous population where queen gene expression abundances in every insect are obtained by balancing the production and degradation terms, leading to a steady state number of queen gene products $n^* = \mu/\delta$. Therefore, a model considering only processes taking place at the molecular level cannot explain the emergence of distinct phenotypes, corresponding to multiple fixed points of the microscopic dynamics, in the steady state. To break the symmetry between the individuals after queen removal, and lead to the emergence of a single queen, a collective process in the population scale is required.

2.3. Symmetry breaking of molecular states

Interactions between insects are crucial in the organisation of colonies of social insects, regulating processes such as the exchange of fluid (trophallaxis) or foraged material between different insects. Due to their central role in the functioning of the society, they have been the subject of numerous studies concerned with their quantification and classification [61, 68–70].

Therefore, to study whether there is a colony-level component in the regulation of reprogramming and phenotypic specialisation we quantified and classified interactions between wasps. Specifically, we analysed over 17 hours of video recordings from 5 nests that had undergone queen removal up to phase D4 (2 nests, 4 phases, $47 \pm \text{SD } 22$ min) and D14 (3 nests, 5 phases, $46 \pm \text{SD } 21$ min). To obtain these recordings, a HD video camera was placed in front of 5 nests during the 2012 fieldwork and video recordings were made for at least 30 min during active hours for each phase of the queen removal experiment. A total of 400 interactions from 80 different individuals were manually classified and quantified in 6 previously defined classes [69, 70]. The six different classes in which interactions were classified are defined as follows:

- Dominance: Dominant behaviour in an aggressive interaction against another insect (Figure 2.6).
- Subdominance: Subordinate behaviour in an aggressive interaction.
- Donor exchange: Donation of foraging material to another insect.
- Receiver exchange: Receipt of foraging material from another insect.

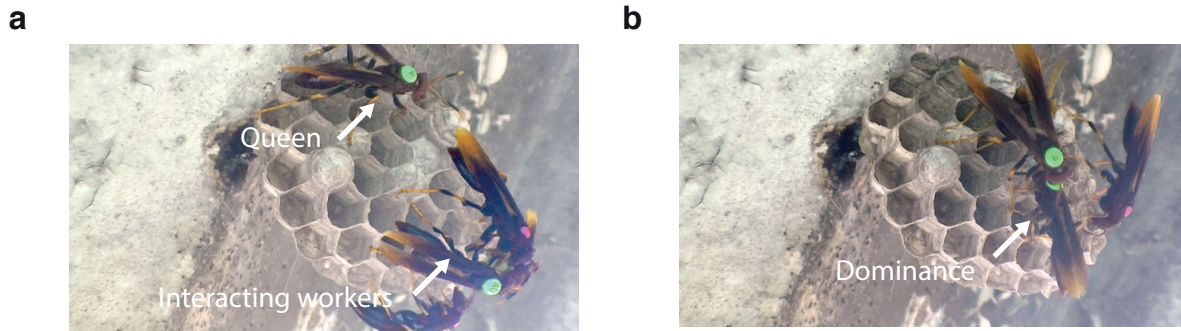


Figure 2.6.: **a)** Example of an aggressive interaction between a queen and a worker. **b)** Example of dominant queen behaviour.

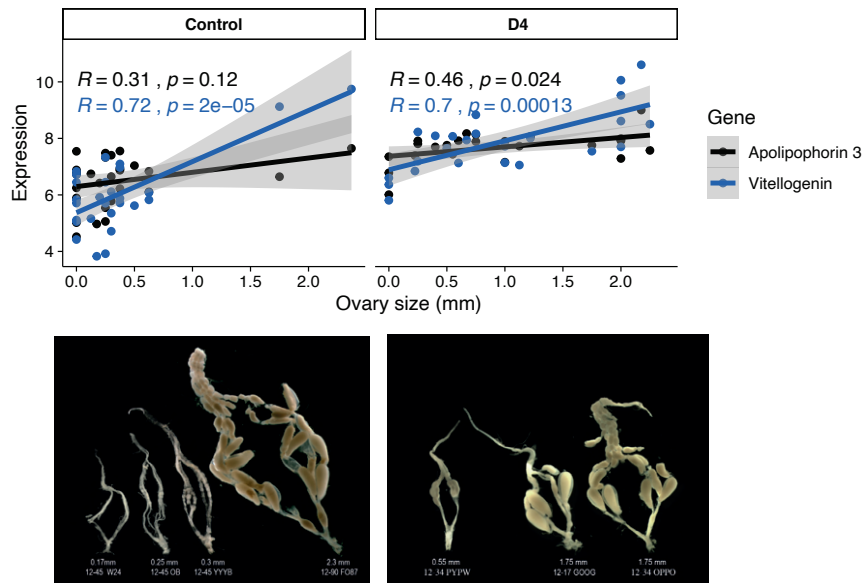


Figure 2.7.: **Top:** Correlation between normalized gene expression and ovary size of two representative queen genes in the control and D4 phases. **Bottom:** Ovary dissections of individuals from the Control and D4 phases.

- Trophallaxis: Trophallactic exchanges of food or fluid between two insects.
- Antennae: Antennae mediated interactions.

We define fighting interactions (Figure 2.6) as those belonging to the dominance and subdominance categories of the above list. Additionally, we refer to donor and receiver exchange interactions as foraging interactions and to trophallactic and antennation interactions as neutral.

In order to correlate the behaviour of single insects with their queen gene expression profiles, we used ovary sizes at the time of collection as a proxy for the level of queen gene expression of individual wasps, as, given that queen genes are associated with egg-laying individuals, the expression of representative queen genes positively correlates with ovary size (Figure 2.7). To quantify the total amount of interactions in a nest we defined the interaction rate per

individual as

$$I_{\text{rate}} = \frac{1}{T} \sum_i I_i, \quad (2.3)$$

where I_i is the number of interactions of type i and T is the total time the individual was present in the nest.

Figure 2.8a shows that there is an increase in the total interaction rate during reprogramming largely due to a significant increase in the number of fighting interactions during the initial stages of the reprogramming process.

In the following we will focus on how the rate of fighting interactions and the likelihood of being dominant depend on the ovary sizes of the individuals involved in aggressive interactions. With this in mind, we defined two measures for each individual (Equation 2.4, Equation 2.5) that we termed fight index and dominance index to quantify, respectively, the fraction of fighting and dominant interactions per unit of time,

$$I_{\text{fight}} = \frac{1}{T} \frac{D + S}{\sum_i I_i}, \quad (2.4)$$

$$I_{\text{dominance}} = \frac{1}{T} \frac{D}{D + S}, \quad (2.5)$$

where as previously $\sum I_i$ denotes the sum over all interaction types, D and S are the number of dominant/subdominant interactions and T is the total time the individual was present in the nest.

We found that in Control nests individuals with developed ovaries engage in fight more often than individuals with smaller eggs and successfully dominate their nestmates (Figure 2.8b-c). After queen removal, the asymmetry between individuals is broken and all individuals, irrespective of their ovary size, interact aggressively, in accordance with the collective up-regulation of the queen genes after queen removal (Figure 2.4b). This asymmetry is restored in the late phases of the reprogramming process (D14). The results show that ovary size positively correlates with the interaction rate and the probability of being dominant, indicating a relationship between queen gene expression and interaction rate. In the remainder of this section we will devise a minimal functional form for the interaction kernel governing the interactions between individual insects based on these experimental observations.

In order to build a physical model, we propose that these interactions have a repressive effect in gene expression, in turn breaking the symmetry between individuals after queen removal. This hypothesis is consistent with experimental observations of workers' ovary sizes regressing after being introduced in a nest with an established dominant queen [60]. In the following, motivated by the quantification of the interactions (Figure 2.8), we derive a minimal

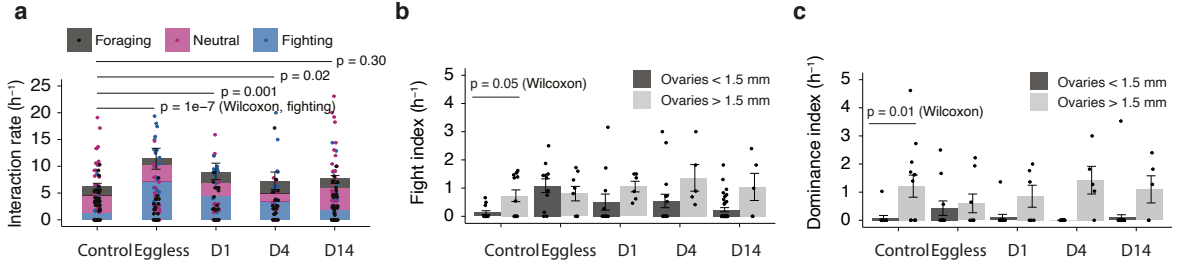


Figure 2.8.: **a)** Total interaction rate separated by type across the different experimental phases **b)** Fight index (Equation 2.4) across the different experimental phases. **c)** Dominance index (Equation 2.5) across the different experimental phases. In all panels bars depict mean \pm SEM, dots represent individuals and p-values were calculated from a Wilcoxon signed-rank test between the corresponding conditions.

interaction kernel that is consistent with the experimental findings and describes how two given insects interact by determining the interaction rate and the dominance probability.

First, we start by decomposing the rate with which individual i is subject to a subdominant interaction with individual j , K_{ij} , as the total interaction rate between both, a_{ij} , times the conditional probability b_{ij} that individual i is subdominant in such an interaction, $K_{ij} = a_{ij}b_{ij}$. To calculate a_{ij} we note that according to our video recordings, the interaction rate of an individual insect increases with ovary size and therefore, with the expression level of queen genes, $a_i \equiv \sum_{j \neq i} a_{ij} \propto n_i$. The pairwise interaction rate therefore is proportional to the probability that both individuals interact in a given time interval, $a_i a_j$, times the probability that this interaction involves individuals i and j , $2/[N(N-1)]$, $a_{ij} = 2a_i a_j/[N(N-1)] \propto 2n_i n_j/[N(N-1)]$, such that we set $a_{ij} = \omega n_i n_j$ where ω is proportional to $2/[N(N-1)]$.

To derive the conditional probability that individual i is subdominant, b_{ij} , we resort to previous work showing that the outcome of an interaction is strongly determined by the concentrations of Insect Juvenile Hormone (JH), a hormone crucial for ovary development [71] that is strongly associated with reproductive phenotypes [72]. Indeed, in our video recording of control nests we found that, in an interaction between two individuals the subdominant one is with high probability the one with the smaller ovaries (Figure 2.8c). Given that ovary size is a proxy for queen gene expression, we take the conditional probability of an individual being subdominant to depend on the gene expression difference between two interacting individuals, $b_{ij} = b(n_j - n_i)$. If the individual with lower queen gene expression is always subdominant, the conditional probability of i being the subdominant individual in an interaction between i and j can be written as

$$b(n_j - n_i) = \Theta(n_j - n_i). \quad (2.6)$$

The total rate of subdominant interactions that individual i receives then takes the form

$$\omega \sum_{j \neq i} K(n_i, n_j), \quad (2.7)$$

with the interaction kernel $K(n_i, n_j)$ defined as

$$K(n_i, n_j) = n_i n_j \Theta(n_j - n_i). \quad (2.8)$$

In order to derive the above form of the interaction kernel we have assumed that individual insects can accurately measure gene expression levels of the insect they are interacting with. A more realistic approach is to take into account that sensing of gene expression levels is associated with an uncertainty. If this uncertainty is distributed following a normal distribution with zero mean and variance σ^2 , the probability of individual i being subdominant in an interaction with individual j is

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\eta^2/2\sigma^2} \theta(n_j - n_i + \eta) d\eta = \frac{1}{2} \left[1 + \text{Erf} \left(\frac{n_j - n_i}{\sqrt{2\sigma^2}} \right) \right], \quad (2.9)$$

with the error function defined as $\text{Erf}(x) = \int_0^x e^{-y^2} dy$. Therefore, the interaction kernel for noise-sensitive individuals reads

$$K(n_i, n_j) = \frac{n_i n_j}{2} \left[1 + \text{Erf} \left(\frac{n_j - n_i}{\sqrt{2\sigma^2}} \right) \right], \quad (2.10)$$

where σ^2 determines how precisely an individual can measure gene expression of other individuals. The above expression is difficult to manipulate both numerically and analytically due to the presence of the error function. It is, however, well approximated by another sigmoidal function

$$K(n_i, n_j) = n_i n_j \frac{e^{-\lambda(n_i - n_j)}}{1 + e^{-\lambda(n_i - n_j)}}, \quad (2.11)$$

with $\lambda = 2\sigma^{-1}$ denoting the sensing sensitivity.

Finally, once we have specified how individual insects interact, it remains to specify the effect that these interactions have on queen gene expression. We model the effect of interactions on gene expression through a binary variable $q_i \in \{0, 1\}$, representing the activation state of the queen genes. The dynamics of this variable are determined by the rate of subdominant interactions, as we will detail in the following section where we combine the molecular and population level processes described in the last two sections to obtain a mathematical description of the dynamics of the system in the form of a master equation describing the

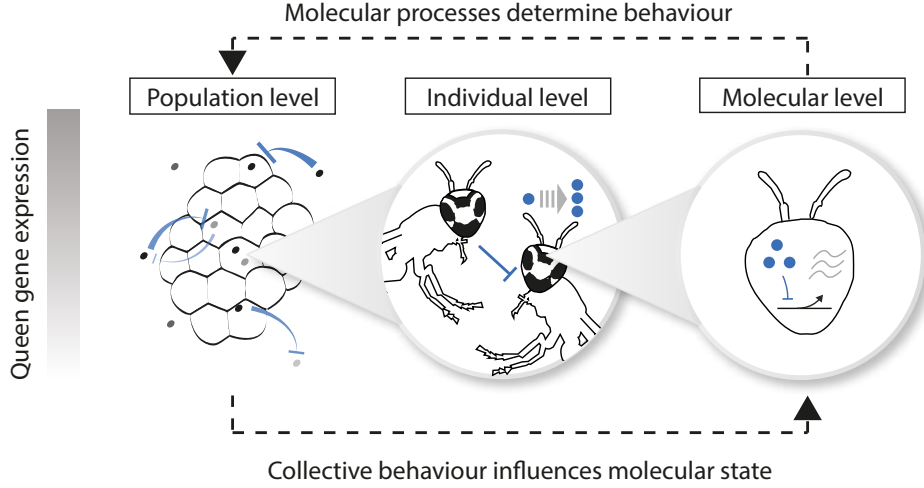


Figure 2.9.: Cartoon representation of the model.

evolution of the joint probability of observing a certain nest composition $P(\{n_k, q_k\})$.

2.4. Master equation description of the multiscale dynamics

Motivated by the experimental observations, we set out to define the minimal model consistent with the experimental findings. Schematically our model is illustrated by Figure 2.9. We consider a nest composed of $N + 1$ individuals described by two degrees of freedom: the number of queen gene products, n_k , and the presence of queen gene repressive factors, q_k . Taken together, the stochastic dynamics describing the evolution of the nest composition $P(\{n_k, q_k\})$ can be cast in the form of the following non-Markovian master equation

$$\begin{aligned} \frac{d}{dt} P(\{n_k, q_k\}) = & \sum_{i=1}^{N+1} \left\{ \mu(1 - q_i) [P(\{n_i - 1, q_i\}) - P(\{n_i, q_i\})] \right. \\ & + \delta [(n_i + 1)P(\{n_i + 1, q_i\}) - n_i P(\{n_i, q_i\})] \\ & \left. + \Gamma(t_i^{\text{int}}) P(\{n_i, 1\})(1 - 2q_i) + \omega \sum_{j \neq i} K_{ij} P(\{n_i, 0\})(2q_i - 1) \right\}, \end{aligned} \quad (2.12)$$

where the terms in the first line account for the stochastic production of queen gene products with rate μ , the second line accounts for the poissonian degradation of these molecules with rate δ and the final one accounts for the dynamics of the queen gene repressor molecules. In particular, subdominant interactions with rate proportional to ωK_{ij} correspond to the transition $q_j : 0 \rightarrow 1$, where K_{ij} is the interaction kernel representing the probability of individual i interacting in a dominant manner with individual j . On the other hand, queen gene repressor molecules are stochastically degraded with a rate proportional to $\Gamma(t_i^{\text{int}})$, where Γ is a distribution that depends on the time elapsed since the last subdominant interaction.

In the following, for simplicity we will consider the deterministic limit where queen gene repressor molecules degrade instantly after a time t_{decay} . In this case, $\Gamma(t) \propto \delta(t - t_{\text{decay}})$.

Finally, to obtain a more compact description of the dynamics, we appropriately rescale time and queen gene product concentrations obtaining the non-dimensional master equation

$$\begin{aligned} \frac{d}{d\tau} P(\{r_k, q_k\}) = & \sum_{i=1}^{N+1} \left\{ (1 - q_i) [P(\{r_i - \epsilon, q_i\}) - P(\{r_i, q_i\})] \right. \\ & + [(r_i + \epsilon)P(\{r_i + \epsilon, q_i\}) - r_i P(\{r_i, q_i\})] \\ & + \Gamma(\tau_i^{\text{int}}) P(\{r_i, 1\})(1 - 2q_i) \\ & \left. + \alpha \frac{\mu}{\delta} \sum_{j \neq i} K(r_i, r_j) P(\{r_i, 0\})(2q_i - 1) \right\}, \end{aligned} \quad (2.13)$$

where $\alpha = \omega/\delta$ is the interaction rate in units of the degradation time, the dimensionless time $\tau = \mu \cdot t$, rescaled queen gene expression levels $r_k = \delta n_k/\mu$ and $\epsilon = \delta/\mu$. Equation 2.13 describes the non-dimensional dynamics of the joint probability distribution representing a particular nest composition, which in turn can be simulated using stochastic sampling algorithms [73] and used to predict the experimental observations. In the next section we will numerically study the reprogramming dynamics of Equation 2.13 starting from a queenless initial population and explore the phase space of the system as a function of the interaction parameters.

2.5. Phase diagram of the system

Equation 2.13 provides the starting point to explore the behaviour of the system. We ran 1000 stochastic simulations of the master equation using Gillespie's algorithm [73] and averaged the results to obtain the distribution of queen gene expression $P(r)$ across simulations. For suitable combinations of the parameters, we found that the distribution of queen gene expression in the system reaches a bimodal steady state starting from unimodal initial conditions (Figure 2.10a). The dynamics are initially characterized by a brief regime where all individuals upregulate the queen genes, followed by the splitting of the distribution of queen gene expression across the population, giving rise to a bimodal distribution that slowly converges to the steady state where the two modes correspond, respectively, to the worker and queen phenotypes.

In order to systematically quantify the effect that interaction parameters have on the steady state of the dynamics we averaged the results of 50 stochastic simulations of a nest of $N = 10$ individuals starting from a queenless initial configuration, where gene expression across the

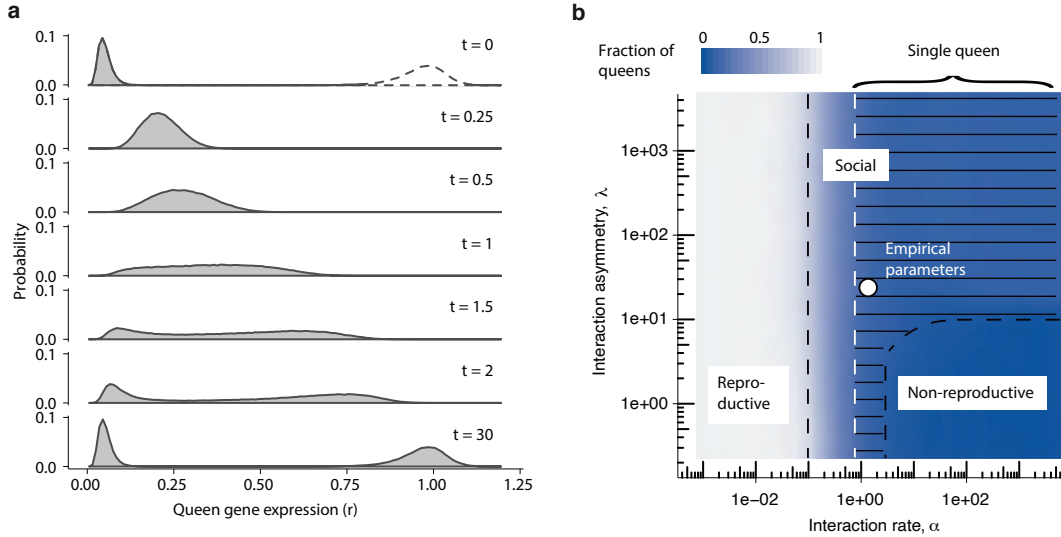


Figure 2.10.: **a)** Evolution of the probability density of queen gene expression across the population after queen removal obtained by stochastic simulations for $\alpha = 0.2$, $\lambda = 10$. **b)** Phase diagram as a function of the interaction rate α and the interaction asymmetry λ .

population is distributed according to a Gaussian distribution of mean $\mu = 100$ and standard deviation $\sigma = 15$, for 900 combinations of the parameters (α, λ) logarithmically distributed in the ranges $\alpha \in [10^{-5}, 10^2]$ and $\lambda \in [10^{-3/2}, 10^3]$, while keeping all other parameters fixed at biologically plausible values (the values of the remaining parameters are given in Appendix A). The phase diagram of the system as a function of the interaction rate α and the interaction asymmetry λ , where the fraction of queens (defined as individuals with a gene expression value larger than 80 % of the steady state value) plays the role of an order parameter, is presented in Figure 2.10b.

The system exhibits three distinct regimes as a function of the interaction parameters. First, for $\alpha \ll 1$ we find a regime that we term 'reproductive'. In this region of the phase space interactions between individuals are rare compared to the typical timescales of the molecular processes required to upregulate the queen genes. Therefore, as interactions do not balance the intrinsic molecular dynamics, we find that starting from an initial queenless configuration the steady state of the system corresponds to a population composed of exclusively queen-like individuals. On the other hand, in the region of the phase space characterised by large values of α and small values of λ , interactions happen on a faster timescale than the molecular processes but the subdominant individual is randomly chosen. In this regime that we term 'non-reproductive' the effect of interactions prevents the upregulation of the queen genes in individual wasps, meaning that the steady state corresponds to a distribution where all individuals remain workers. Finally, in between these two regimes we find a large region

of the phase space where the steady state corresponds to a bimodal distribution of queen gene expression across the population with the two modes of this distribution representing the worker and queen phenotypes. We refer to this region of the phase space as the 'social' regime. The relative weights of the two modes are controlled by the value of the interaction parameters α and λ . Moreover, for a large section of this regime our model predicts the existence of a single queen in the steady state, corresponding to the biologically relevant section of the phase diagram and indicating that the system is robust against fluctuations in the interaction parameters, not requiring any careful parameter tuning to operate in the social regime. Our theoretical work in Section 2.7 reveals that the transition between the specialised and the reproductive regimes corresponds to a saddle node bifurcation where the population structure plays the role of control parameter.

In order to compare our theoretical results with the experimental data, we estimated the interaction parameters α and λ from the empirical data presented in Figure 2.8. First, to estimate the interaction rate, we calculated the number of aggressive interactions per individual in a time interval of one day obtaining an estimate of 3.7 interactions per day. The rescaled interaction rate α corresponds to this quantity measured in the units defined by the degradation timescale $\delta^{-1} \approx 12\text{h}$. Therefore, we obtain $\alpha_{\text{est}} \approx 1.85$.

Additionally, we counted the number of subdominant queen interactions to obtain a lower bound for the interaction asymmetry parameter λ . Out of 17 interactions involving queens in the control and late-commitment phases, we observed 0 interactions where the queen was subdominant. Therefore, the maximum likelihood estimate for the error rate using the beta-distribution as a (conjugate) prior is $1/17$. Using the definition of the interaction kernel we find an analytical expression for the error rate,

$$2 \int_{-\infty}^0 d\Delta r (1 + \exp(-\lambda\Delta r))^{-1} = 2 \ln 2/\lambda. \quad (2.14)$$

Comparing with the maximum likelihood estimate we find a lower bound for the asymmetry parameter $\lambda_{\text{est}} \approx 24$ that we take to be the empirical estimate of λ .

In conclusion, our estimate of the interaction parameters $(\alpha_{\text{est}}, \lambda_{\text{est}})$ is given by $(\alpha_{\text{est}}, \lambda_{\text{est}}) \approx (1.85, 24)$. The empirical estimates of the interaction parameters are located inside the single queen region of the phase diagram (Figure 2.10b) for a nest of size $N = 10$, indicating that our model correctly predicts the presence of a single queen in the steady state nest for a biologically relevant combination of parameters. Furthermore, the size of the single-queen region of the phase diagram reflects that the model is robust against small changes in the interaction parameters. Therefore, the presence of a single queen in the steady-state is an

emergent property of the system arising from the interplay between antagonistic processes across several scales of biological organisation. In the following sections we study how this interplay also gives rise to the plastic response of the colony after queen removal.

2.6. Mean-field master equation description of the nest dynamics

Although the master equation approach derived in the previous section (Equation 2.13) can accurately reproduce the reprogramming dynamics, it is nevertheless too complex to provide an intuitive understanding of the physical mechanisms underlying specialisation and plasticity in *Polistes*, as its high dimensionality and non-Markovianity render it unsuitable for analytical treatment. In the following we derive an approximation of Equation 2.13 that still captures correctly the structure of the phase space of the system while at the same time being more amenable to analytical treatment. We refer to the result of this approximation as the mean-field master equation. Thereafter, we will employ this continuum description in Section 2.7 to explore the structure of the phase space as a function of the individual and collective degrees of freedom.

In order to derive the mean-field master equation we start again from the master equation describing the evolution of the joint probability distribution $P(\{r_k, q_k\}, t)$ (Equation 2.13):

$$\begin{aligned} \frac{d}{d\tau} P(\{r_k, q_k\}) = & \sum_{i=1}^{N+1} \left\{ (1 - q_i) [P(\{r_i - \epsilon, q_i\}) - P(\{r_i, q_i\})] \right. \\ & + [(r_i + \epsilon)P(\{r_i + \epsilon, q_i\}) - r_i P(\{r_i, q_i\})] \\ & + \Gamma(\tau_i^{\text{int}}) P(\{r_i, 1\}) (1 - 2q_i) \\ & \left. + \alpha \frac{\mu}{\delta} \sum_{j \neq i} K(r_i, r_j) P(\{r_i, 0\}) (2q_i - 1) \right\}. \end{aligned}$$

To begin, we perform a Hartree approximation [74, 75] in order to study the evolution of a single individual, that we refer to as the 'tracer', embedded in a nest with a given composition $P(\{r_i, q_i\}_{i=1}^N)$. In this approach we consider the nest as a 'bath' that determines the fluctuations of the individual tracer dynamics. Mathematically, we write down the equation determining the time-evolution of the probability of finding the tracer in the state (r, q) given

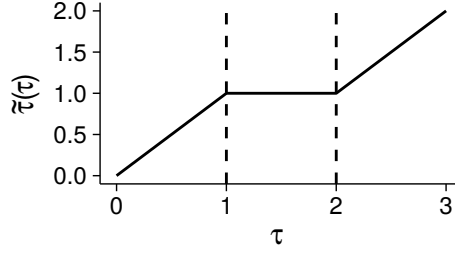


Figure 2.11.: Definition of the trajectory dependent time $\tau(t)$. Dashed lines indicate the beginning and the end of the repression of queen gene expression after a subdominant interaction.

the nest composition, $P(r, q) \equiv P(r, q|\{r_i, q_i\})$. This equation reads

$$\begin{aligned} \frac{d}{d\tau}P(r, q) &= (1 - q) [P(r - \epsilon, q) - P(r, q)] \\ &+ [(r + \epsilon)P(r + \epsilon, q) - rP(r, q)] \\ &+ \Gamma(\tau_{int}^i)P(r, 1)(1 - 2q) + \alpha \frac{\mu}{\delta} \sum_{j=1}^N K(r, r_j)P(r, 0)(2q - 1). \end{aligned} \quad (2.15)$$

To obtain a time evolution equation for the marginal probability $P(r)$ we first need to integrate out the queen gene repressors variable, q . To this end, we formally define a trajectory dependent time (Figure 2.11) as

$$\tilde{\tau}(t) = \int_0^\tau d\tau' \prod_{i \in \mathcal{I}} f(\tau_i - \tau'), \quad (2.16)$$

where $\{\tau_i\}_{i \in \mathcal{I}}$ is the set of times when an interaction, and hence a $q : 0 \rightarrow 1$ transition, takes place, and the function $f(\tau_i - \tau)$ is defined as

$$f(\tau_i - \tau) = \begin{cases} 0 & \tau \in [\tau_i, \tau_i + \tau_{\text{decay}}], \\ 1 & \text{otherwise.} \end{cases} \quad (2.17)$$

From this definition it follows that $\tilde{\tau}$ increases in the same manner as τ if and only if $q = 0$ and is constant otherwise. The evolution of the system in the new time coordinate $\tilde{\tau}$ thus coincides with the evolution with respect to τ when $q = 0$ and its effect is collapsed to a single time point when $q = 1$. This is equivalent to considering dynamics where $q = 0$ at all times and where a suitably chosen number, M , of queen gene products are instantaneously degraded at times $\tau \in \{\tau_i\}_{i \in \mathcal{I}}$. M is the typical number of molecules degraded in a time interval of length τ_{decay} , $M \approx r[1 - \exp(-\tau_{\text{decay}})]$. The master equation describing such dynamics takes

the form

$$\begin{aligned}
 \frac{d}{d\bar{\tau}}P(r) &= [P(r - \epsilon) - P(r)] \\
 &+ [(\epsilon + r)P(r + \epsilon) - rP(r)] \\
 &+ \alpha \frac{\mu}{\delta} \sum_{j=1}^N [K(r + M, r_j)P(r + M) - K(r, r_j)P(r)] .
 \end{aligned} \tag{2.18}$$

Next, we consider time scales much longer than typical interaction times. To derive a continuum description, we rescale the persistence time of the repressive factors τ_{decay} in such a way that the average number of degraded queen gene products after an interaction $M(\tau_{\text{decay}})$ is, on average, equal to a queen gene product unit, ϵ . Simultaneously, we keep the total effect of repressive interactions in a long time interval $[\tau_1, \tau_2]$ on the concentration of queen gene products, $\int_{\tau_1}^{\tau_2} d\bar{\tau}' M(\bar{\tau}'_{\text{decay}}) \prod_{i \in \mathcal{I}} \delta(\bar{\tau}'_i - \bar{\tau}')$, invariant. To this end, the above constraint dictates a corresponding rescaling of the interaction rate, α . Intuitively, such a coarse-graining operation corresponds to a homogeneous distribution of interaction events in the time domain for sufficiently long time scales. As the total effect of interactions on long time scales remains unchanged the structure of the phase portrait remains unchanged as well. With this, we obtain

$$\begin{aligned}
 \frac{d}{d\bar{\tau}}P(r) &= [P(r - \epsilon) - P(r)] \\
 &+ [(r + \epsilon)P(r + \epsilon) - rP(r)] \\
 &+ \alpha' \sum_{j=1}^N [K(r + \epsilon, r_j)P(r + \epsilon) - K(r, r_j)P(r)] ,
 \end{aligned} \tag{2.19}$$

where $\alpha' \approx \alpha\mu M/\delta$ is the rescaled interaction rate. The mean-field master equation is then obtained by setting $x = r/\Omega$, and performing a Kramers-Moyal expansion to the lowest order [76]. Retaining the symbol r to represent queen gene expression for the sake of simplicity, we find

$$\partial_{\bar{\tau}}r = \tilde{\alpha}_1(x) = 1 - r - \tilde{\alpha} \sum_{j=1}^N K(r, r_j) , \tag{2.20}$$

where

$$\tilde{\alpha}_1(r) = \Omega^{-1} \int_{-\infty}^{\infty} dr' (r' - r) W(r'|r) , \tag{2.21}$$

is the first jump moment and $\tilde{\alpha} = \alpha'/\Omega^2$. We henceforth refer to the dual version of Equation 2.20 describing the evolution of the probability density $P(r, \bar{\tau})$ as the mean field master

equation,

$$\partial_{\bar{t}}P(r) + \partial_r [(1-r)P(r)] = \tilde{\alpha}\partial_r \left(P(r) \sum_{j=1}^N K(r, r_j) \right). \quad (2.22)$$

Equation 2.22 describes the evolution of the queen gene expression of a tracer individual embedded in a finite population of N individuals with queen gene expression levels $\{r_j\}_{j=1}^N$. In the next section, we neglect fluctuations arising from the finite size of the population by taking the limit $N \rightarrow \infty$ and derive a phase portrait description of the co-evolution of the molecular and population degrees of freedom from the resulting evolution equation.

2.7. Continuum limit of the mean-field master equation and phase portrait

In the previous section we have derived a partial differential equation describing the evolution of the probability distribution of the queen gene expression levels in a single individual embedded in a finite nest. In order to obtain a mean-field master equation suitable for analytical treatment, we further simplify our description by taking the continuum limit of Equation 2.20 on the number of individuals, $N \rightarrow \infty$, obtaining the following equation for the time evolution of queen gene expression levels in the 'tracer' wasp,

$$\partial_t r = 1 - r - \tilde{\alpha} \sum_{i \neq j} K(r, r_j) = 1 - r - \tilde{\alpha} \int_0^\infty \sum_{j=1}^N K(r, r') \delta(r' - r_j), \quad (2.23)$$

$$1 - r - \tilde{\alpha} \int_0^\infty \sum_{j=1}^N K(r, r') \delta(r' - r_j) \xrightarrow{N \rightarrow \infty} 1 - r - \tilde{\alpha} \int_0^\infty K(r, r') f(r') dr', \quad (2.24)$$

where the empirical distribution function $f(r') \equiv \sum_j \delta(r_j - r')$ represents the fraction of individuals in the nest having queen gene expression between r and $r + dr$, as well as the time evolution of the population composition,

$$\partial_{\bar{t}}f(r) + \partial_r [(1-r)f(r)] = \tilde{\alpha}\partial_r \left(f(r) \int_0^\infty K(r, r') f(r') dr' \right). \quad (2.25)$$

These two equations represent the mean-field description of Equation 2.13 valid in the limit of large populations and time scales, i.e. in the steady state. As we will discuss below, while Equation 2.24 and Equation 2.25 are not suitable for quantitatively describing the reprogramming dynamics they nevertheless are capable of providing mathematical insight into the mechanisms underlying the regulation of specialisation and plasticity. Interestingly, Equation 2.25 is conceptually similar to equations obtained in other biological contexts, such as

quorum-sensing bacteria [77] or Mitogen competition by stem cells [78]. In this context, it is also worth to note that in spatially structured systems specialisation can be achieved by spatially separating different phenotypes, such as via the Turing mechanism, spinodal decomposition, lateral inhibition or via external signalling gradients [18, 79]. If the spatially homogeneous state is unstable such systems are naturally 'plastic'.

The dynamics described by Equation 2.24 and Equation 2.25 represent production and degradation of queen gene products on the molecular scale coupled through a collision-like functional that accounts for the effect of repressive interactions. The steady state of these equations is reached when the molecular dynamics are balanced by the population-level feedback. In the following we will study how such a coupling gives rise to specialisation and plasticity as emergent properties of the system.

To understand the relaxation dynamics to the steady state, and their stability, it is instructive to consider the co-evolution of the molecular scale, represented by the queen gene expression level r , and the population scale, represented by the distribution $f(r, t)$. In this section we will illustrate the results of our analysis by means of a phase portrait of the multi-scale dynamics. Although the limits we take, such as taking the mean-field limit, do not accurately reflect the full biological complexity, our approximations are validated by comparison to simulations of the full stochastic dynamics.

The starting point of our analysis are Equation 2.24 and Equation 2.25, describing the individual and collective dynamics, respectively. Taken together, these equations describe the co-evolution of the queen gene expression level of an individual and the population structure. Stable fixed points of such dynamics represent possible phenotypes in the society, such as queen and workers. We calculate these fixed points from the intersection of the nullclines of the system. These nullclines are given by

$$0 = 1 - r - \tilde{\alpha} \int_{-\infty}^{\infty} K(r, r') f(r') dr' , \quad (2.26)$$

$$\partial_r [(1 - r) f(r, t)] = \tilde{\alpha} \partial_r \left(f(r, t) \int_{-\infty}^{\infty} K(r, r') f(r', t) dr' \right) . \quad (2.27)$$

Equation 2.26 and Equation 2.27 provide the basis for understanding the steady state of the system as a function of the population composition and the fixed points of the individual dynamics. Considering that the population composition is represented by a probability distribution, f , its functional nature complicates intuitive interpretations of the relaxation dynamics. In order to obtain a more intuitive picture, we reduce the system to a two-dimensional system describing the coupled evolution of r and the first moment of the population composition,

$\langle r \rangle$. The starting point of this approximation is Equation 2.24,

$$\partial_{\bar{t}} r = 1 - r - \tilde{\alpha} \int_0^\infty r r' \Theta(r' - r) f(r') dr', \quad (2.28)$$

describing the evolution of the molecular degree of freedom of a tracer individual. The integral in the right hand side of the equation represents the effect of the interactions received by the tracer individual. In the limit of long times compared with the typical interaction time scale, $t \gg \alpha^{-1}$, the effect of interactions can be approximated by the overall effect of interacting with an effective individual with gene expression level $\langle r \rangle$,

$$\partial_{\bar{t}} r \approx 1 - r - \tilde{\alpha} r \langle r \rangle \Theta(\langle r \rangle - r), \quad (2.29)$$

where $\langle r \rangle = \int_0^\infty r f(r, t) dr$. Further, by multiplying Equation 2.25 by r and integrating over r , we obtain the time evolution of the first moment,

$$\partial_{\bar{t}} \langle r \rangle = 1 - \langle r \rangle - \tilde{\alpha} \int_0^\infty dr \int_0^\infty dr' K(r, r') f(r) f(r'), \quad (2.30)$$

where $\int_0^\infty dr \int_0^\infty dr' K(r, r') f(r) f(r')$ is the total interaction rate at time t in the nest. To close the system of equations we approximate $\int_0^\infty dr \int_0^\infty dr' K(r, r') f(r) f(r') \approx \langle r \rangle^2 / 2$ where the factor 2 arises due to double counting of subdominant interactions. Finally, our reduced system of equations reads

$$\partial_{\bar{t}} r \approx 1 - r - \tilde{\alpha} r \langle r \rangle \Theta(\langle r \rangle - r), \quad (2.31)$$

$$\partial_{\bar{t}} \langle r \rangle \approx 1 - \langle r \rangle - \tilde{\alpha} \frac{\langle r \rangle^2}{2}. \quad (2.32)$$

Equation 2.31 and Equation 2.32 form a two-dimensional system of equations that is amenable for a bidimensional graphical representation in the form of a phase portrait (Figure 2.12). The solutions of these equations in the steady state provide the fixed points of the dynamics. For $\tilde{\alpha} \neq 0$, we find in the steady state that

$$\langle r \rangle_0 = \frac{\sqrt{2\tilde{\alpha} + 1} - 1}{\tilde{\alpha}}. \quad (2.33)$$

For $\tilde{\alpha} = 0$ Equation 2.31 and Equation 2.32 admit only one stable solution corresponding to high queen gene expression levels, $r_0 = 1$ (Figure 2.13). On the other hand, for $\tilde{\alpha} > 0$, we find three solutions if

$$\langle r \rangle > \frac{\sqrt{4\tilde{\alpha} + 1} - 1}{2\tilde{\alpha}}, \quad (2.34)$$

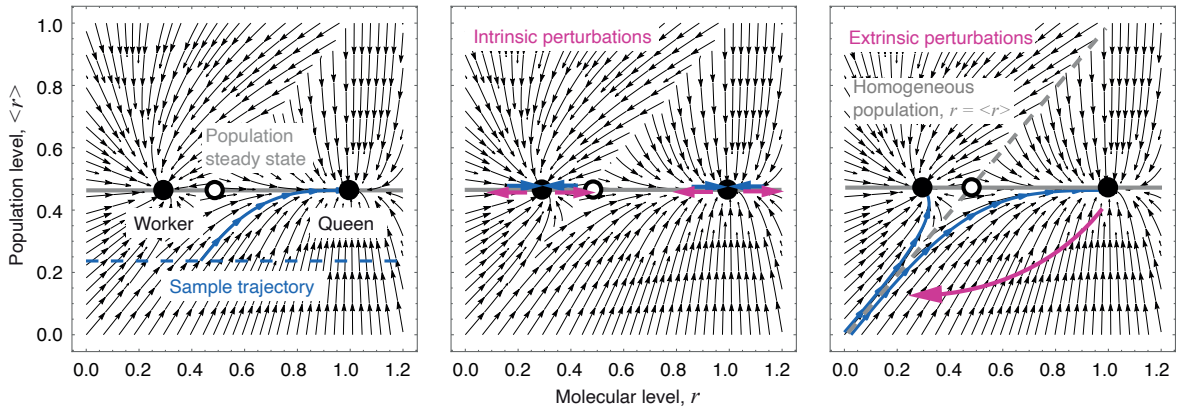


Figure 2.12.: Phase portrait depicting the co-evolution of the population composition, represented by the average queen gene expression across the population $\langle r \rangle$, and the microscopic degree of freedom, represented by the queen gene expression of single insects r . Left: Unperturbed dynamics, center: effect of intrinsic perturbations, right: effect of extrinsic perturbations.

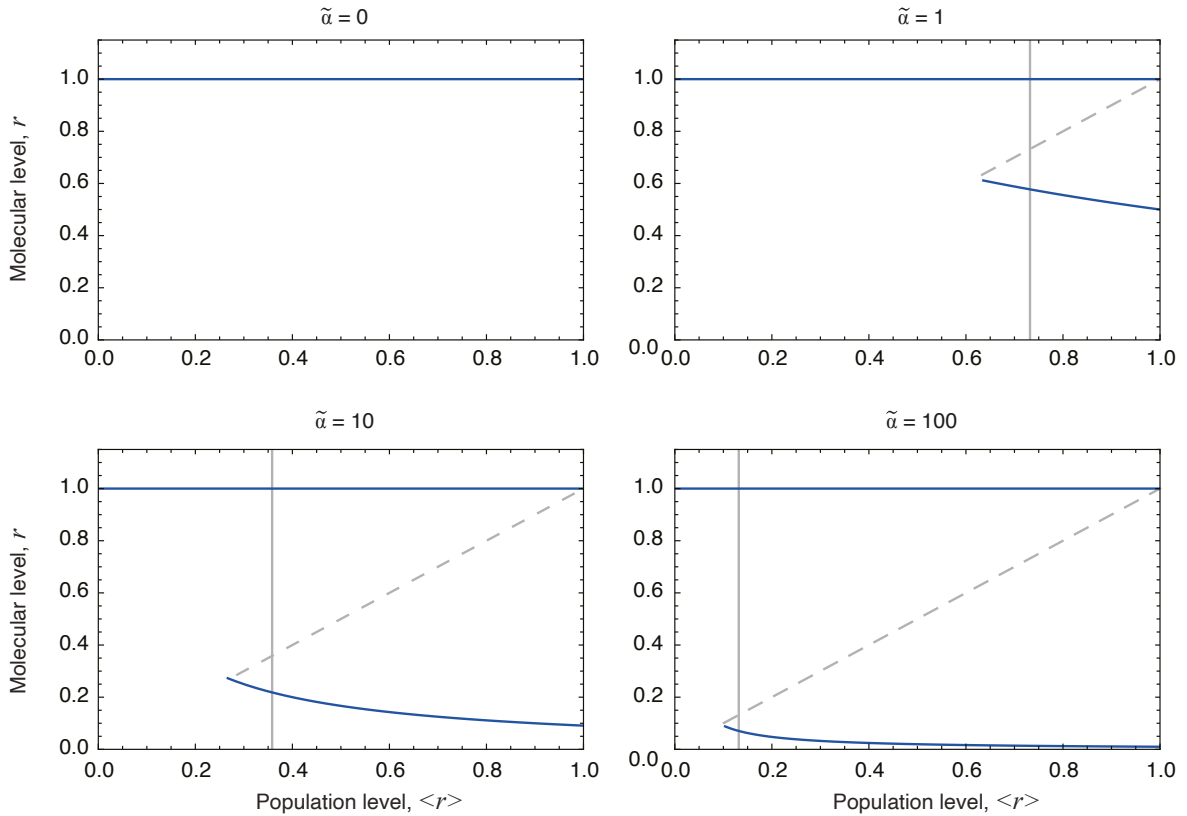


Figure 2.13.: Bifurcation diagram showing fixed points of Equation 2.31 as a function of $\langle r \rangle$ for different values of the interaction rate $\tilde{\alpha}$. Stable branches are represented by solid lines and unstable branches by dashed lines. Vertical lines denote the nullclines of Equation Equation 2.32. For $\tilde{\alpha} > 0$ the dynamics undergo a saddle node bifurcation with the population composition, $\langle r \rangle$, as a bifurcation parameter.

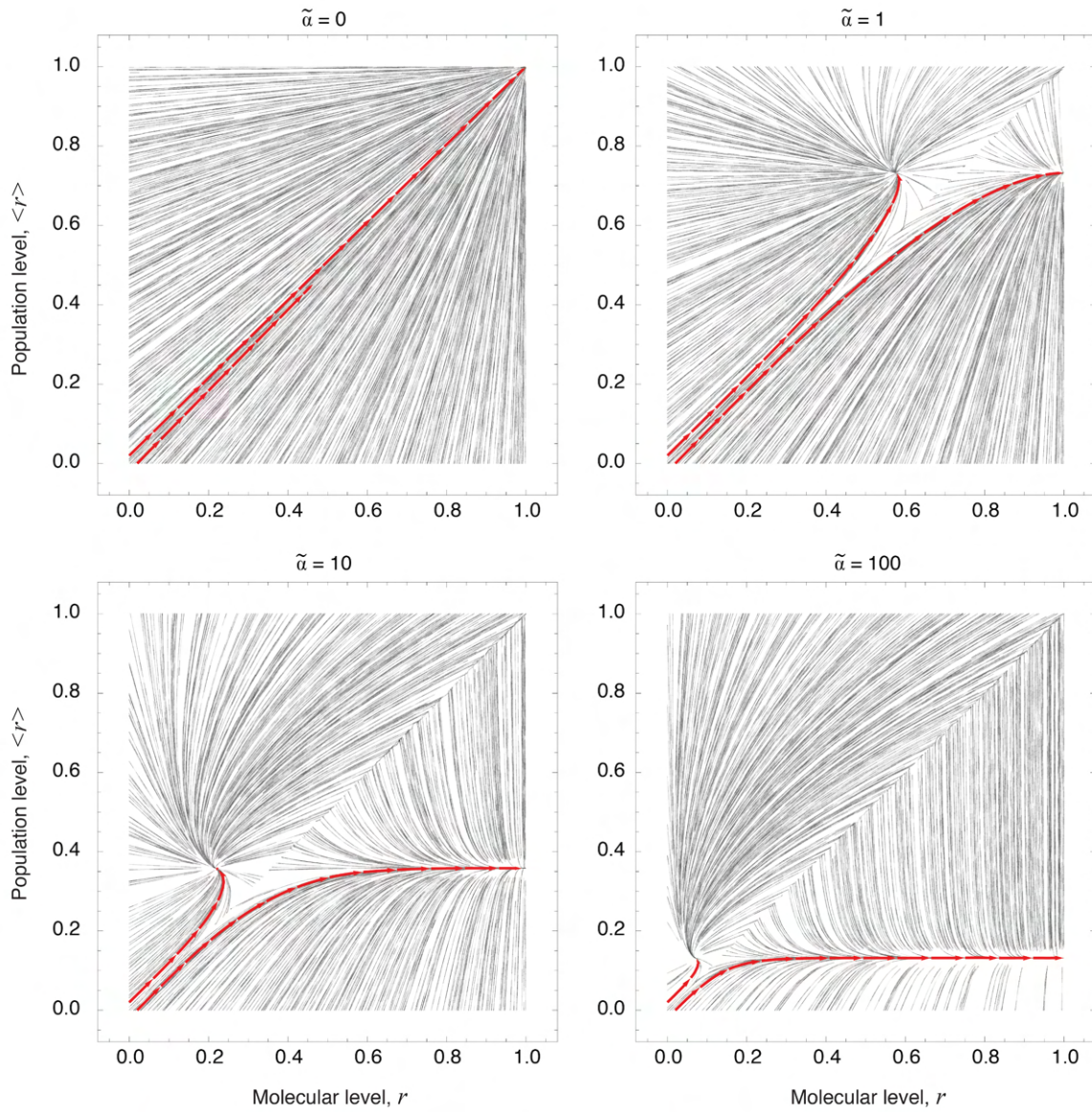


Figure 2.14.: Phase portraits of Equation 2.31 and Equation 2.32 for different values of $\tilde{\alpha}$. Random trajectories are represented by gray lines and two trajectories originating around the point $(0, 0)$ corresponding to a nest with only individuals lacking queen gene expression are highlighted in red.

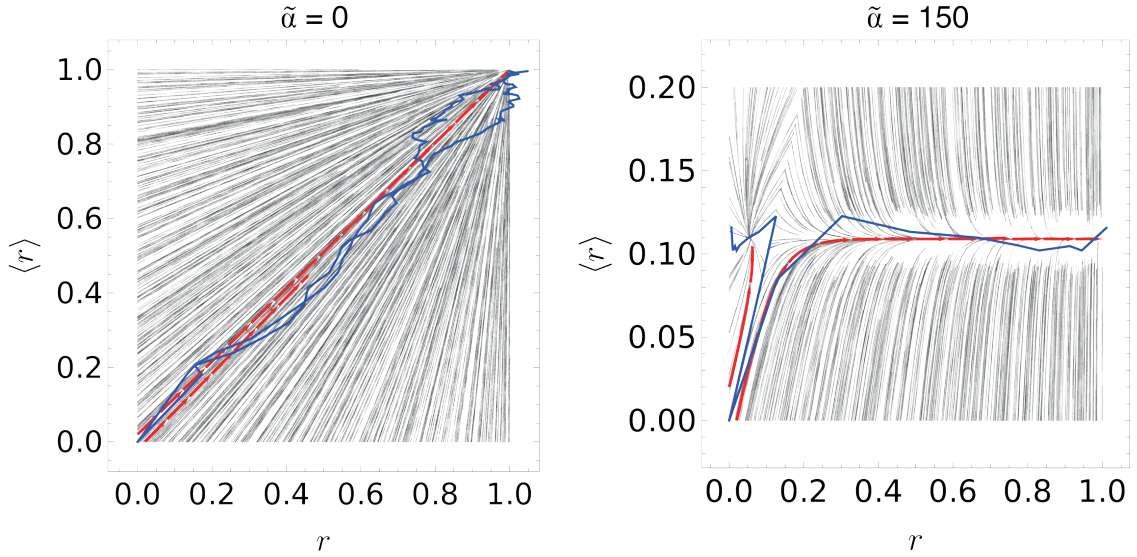


Figure 2.15.: Phase portrait for different values of the interaction rate $\tilde{\alpha}$. Highlighted in red are two example trajectories with similar initial conditions. The result of stochastic simulations is overlaid in blue.

which in the steady state is always fulfilled for $\tilde{\alpha} > 0$. These three solutions comprise two stable branches, at $r_0 = 1$ and $r_2 = 1/(1 + \tilde{\alpha}\langle r \rangle)$ and an unstable branch at $r = \langle r \rangle$. Formally, the system therefore comprises a saddle node bifurcation with the population structure as a bifurcation parameter, as it had been previously observed in other contexts [80]. Intrinsic perturbations (Figure 2.12b), which do not change the value of $\langle r \rangle$, are therefore suppressed by the bistable dynamics in the steady state. Substituting the population steady state, $\langle r \rangle_0$ we obtain as intersections of the nullclines for $\tilde{\alpha} > 0$,

$$\begin{aligned} r'_0 &= 1, \\ r'_1 &= \frac{\sqrt{2\tilde{\alpha} + 1} - 1}{\tilde{\alpha}}, \\ r'_2 &= \frac{1}{\sqrt{2\tilde{\alpha} + 1}}. \end{aligned} \tag{2.35}$$

On the other hand, after an extrinsic perturbation like queen removal (Figure 2.12c) the population structure becomes homogeneous and narrowly distributed around the separatrix defined by $r = \langle r \rangle \ll 1$. From these initial conditions, the dynamics evolve rapidly along the separatrix. On this separatrix fluctuations can drive each individual either to the queen or to the worker attractor, as reflected in the divergence of trajectories with similar initial conditions shown in Figure 2.14.

The flow in the mean-field limit qualitatively represents the flow obtained from stochastic simulations of Equation 2.13 using Gillespie's algorithm [73], as shown in Figure 2.15. There-

fore, while our calculations are strictly only valid in the limit of an infinite population size and on long time scales, Equation 2.31 and Equation 2.32 accurately represent the qualitative structure of the deterministic phase space of Equation 2.13.

2.8. Quantitative prediction of experimental observables

Inspired by the experimental observations, in previous sections we have derived a model (Equation 2.13) that describes the evolution of the joint distribution of queen gene expression and queen gene repressors and shown that it recapitulates some qualitative features of the system as, for instance, reprogramming after queen removal (Figure 2.10a). In this section, we set out to fix the parameters of the model and quantitatively predict several observables of the experimental dynamics.

First, to derive a mathematical description that is less dependent on parameters describing poorly understood molecular processes, such as the dynamics of queen gene repressors, we derive an effective description of the time evolution of the marginal distribution of ovary sizes. To this end, we derived a description based on ovary growth starting from the model derived above. As ovary sizes are a mere downstream effect of the queen gene expression dynamics, such an effective description should be structurally similar to the master equation describing the time evolution of the marginal distribution of queen gene expression levels, Equation 2.18, if the time scales are chosen appropriately.

Finally, starting from this description we computed the contribution of the fighting interactions to the activity of the nest across the different phases of the experiment and compared our theoretical predictions with the computational quantification of the activity from the colony recordings.

2.8.1. Marginal distribution of ovary sizes

Ovary sizes are one of the defining phenotypical features of *Polistes* wasps, acting as a proxy for queen gene expression, whose expression is directly correlated with ovary size (Figure 2.7a). Additionally, only mature eggs, i.e. larger than 1.5 mm in size, can be laid by these insects (Figure 2.7b). Motivated by these observations, and in order to predict the time evolution of the probability of observing ovaries of size $\{o_k\}$ at time t , $P(\{o_k\}, t)$, we first assume that the instantaneous rate of ovary growth is given by a function that depends on the expression level of queen genes, $g(n_k)$, where the function $g(n)$ summarises a cascade of molecular pathways which are not understood in detail. We here make the simplest possible assumption about the functional form of $g(n)$, namely that it depends linearly on gene expression levels and that

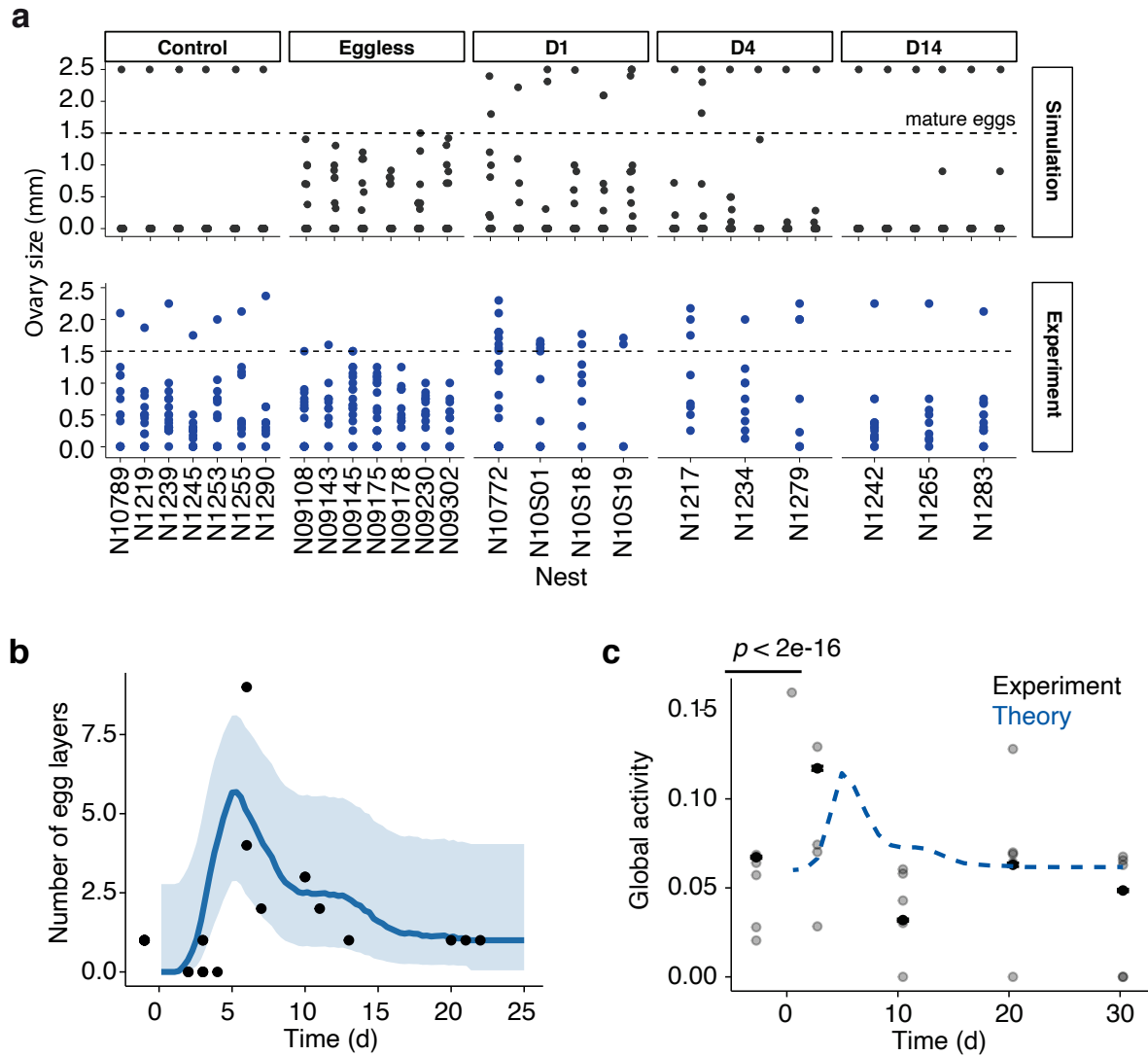


Figure 2.16.: **a)** Top: Theoretical prediction of ovary sizes for the different phases of the experiment obtained from stochastic simulations. Bottom: Experimentally measured ovary sizes for the different experimental phases. **b)** Number of egg-layers for the different phases of the experiment. The solid line corresponds to the theoretical prediction. Shaded area denotes 95% confidence bounds of the theoretical prediction for individual nests **c)** Global activity for the different phases of the experiment. The dashed line corresponds to the theoretical prediction.

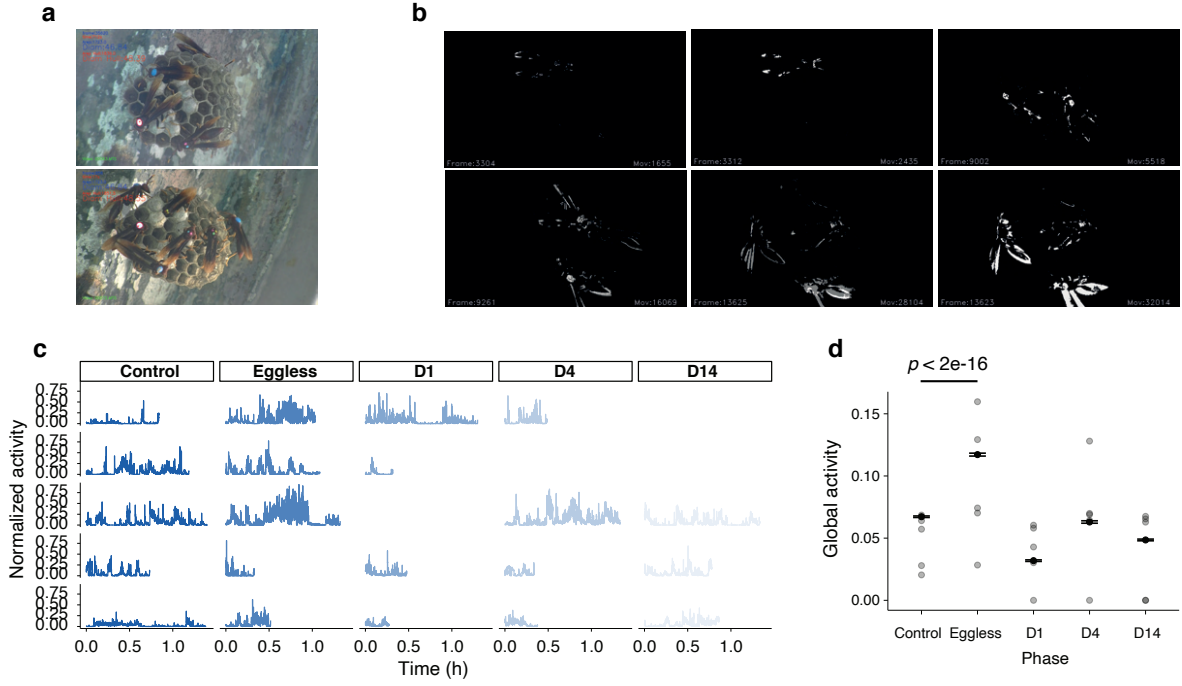


Figure 2.17.: **a)** Example frames of the nest recordings. **b)** Example of an increase in activity detected by pixel changes (in white) between consecutive frames. **c)** Activity over time for individual nests as a function of the recording time. **d)** Average activity over time for the experimental phases (p-value, t-test). Dots represent individual nests. Error bars correspond to SEM.

ovaries grow when the gene expression level exceeds a threshold, n_0 , and shrink otherwise. Therefore, the rate of ovary growth takes the form $g(n) \propto n - n_0$. As eggs are laid once they have reached a mature size, we impose a reflective boundary condition on o_i at a size o_0 which we set to be the maximum ovary size observed in the experiment, $o_0 = 2.5$ mm. With this, switching back to dimensional quantities, the time evolution of the conditional probability of ovary sizes follows a master equation of the form

$$\begin{aligned} \frac{d}{dt} P(\{o_k\}, t | \{n_k, q_k\}) = & g(n_k) \theta(n_k - n_0) [P(\{o_k - 1\}, t | \{n_k, q_k\}) - P(\{o_k\}, t | \{n_k, q_k\})] \\ & - g(n_k) \theta(n_0 - n_k) [P(\{o_k + 1\}, t | \{n_k, q_k\}) - P(\{o_k\}, t | \{n_k, q_k\})] . \end{aligned} \quad (2.36)$$

As in this model the expression level of queen genes is independent of ovary size, the joint probability $P(\{n_k, o_k\}, t)$ factorizes as $P(\{n_k, o_k\}, t) = P(\{o_k\} | \{n_k\}, t) P(\{n_k\}, t)$. Therefore,

the dynamics of the joint probability of $\{n_k, o_k, q_k\}$ are described by

$$\begin{aligned}
 \frac{d}{dt}P(\{n_k, o_k, q_k\}) = & \sum_{i=1}^{N+1} \left\{ \mu(1 - q_i) [P(\{n_i - 1, o_i, q_i\}) - P(\{n_i, o_i, q_i\})] \right. \\
 & + \delta[(n_i + 1)P(\{n_i + 1, o_i, q_i\}) - n_i P(\{n_i, o_i, q_i\})] \\
 & + g(n_i)\theta(n_i - n_0) [P(\{n_i, o_i - 1, q_i\}) - P(\{n_i, o_i, q_i\})] \\
 & - g(n_i)\theta(n_0 - n_i) [P(\{n_i, o_i + 1, q_i\}) - P(\{n_i, o_i, q_i\})] \\
 & + \Gamma(t_i^{\text{int}})P(\{n_i, o_i, 1\})(1 - 2q_i) \\
 & \left. + \omega \sum_{j \neq i} K(n_i, n_j)P(\{n_i, o_i, 0\})(2q_i - 1) \right\}, \tag{2.37}
 \end{aligned}$$

with $g(n)$ denoting the rate of ovary growth for a given gene expression level n .

With the aim of comparing our simulation results to the experimental ovary dissection data we first integrate out gene expression, yielding an equation describing the evolution of ovary sizes. If the persistence time of the repressive effect of interactions plus the typical production time of gene products is smaller than the typical time between two interactions, individuals alternate periods of growing and shrinking of their ovaries with the duration of these periods determined by the ratio between the typical interaction and persistence times plus the queen gene production times. Taking this limit, the ovary growth rate only depends on whether queen gene expression is above or below the threshold n_0 . For that purpose we define a binary random variable,

$$s_i = \Theta(n_i - n_0), \tag{2.38}$$

such that the ovary growth rate is proportional to $2s_i - 1$.

The time evolution of the random variable s_i is linked to the dynamics of the queen gene expression level n_i , giving rise to an implicit time delay in the ovary equation. Specifically, following an interaction, an individual with $s_i = 1$ needs a time t_{off} to degrade enough gene products and activate the pathways responsible for flipping the ovary growth state, $s_i = 0$. On the other hand, if an individual with $s_i = 0$ does not engage in a subdominant interaction during a time $t_{\text{on}} \approx t_{\text{off}} + t_{\text{per}}$, given by the sum of the persistence time of queen gene repressors and the time needed to express queen genes beyond a level n_0 and activate pathways related to reproduction, it will again flip the ovary growth state, $s_i = 1$.

Finally, in order to obtain a description of the joint evolution of ovary sizes and growth states $P(\{o_k, s_k\})$ we still need to specify how interaction probabilities depend on ovary sizes. In the long time limit the ovary size of a given individual is determined by the sign of the difference $T_{s_i=1} - T_{s_i=0}$ where $T_{s_i=j} = \int_0^t dt' \theta(s_i - j)$ is the time that the ovary growth rate

spends in the state j . In this limit, the instantaneous value of gene expression n at time t is a good indicator of the sign of $T_{s_i=1} - T_{s_i=0}$, and hence of ovary growth. Therefore, we can rewrite the interaction kernel in terms of ovary size and obtain the effective master equation governing the evolution of the marginal distribution of ovary sizes,

$$\begin{aligned} \frac{d}{dt}P(\{o_k, s_k\}) = \sum_{i=1}^{N+1} \left\{ g(s_i) [P(\{o_i - 1, s_i\}) - P(\{o_i, s_i\})] \right. \\ \left. + \delta(t_i^{\text{int}} - t_{\text{off}})P(\{o_i, 1\})(1 - 2s_i) + \delta(t_i^{\text{int}} - t_{\text{on}})P(\{o_i, 0\})(2s_i - 1) \right\}, \end{aligned} \quad (2.39)$$

where t_i^{int} is, as before, the time elapsed since the last subdominant interaction of individual i . The dynamics of ovary growth state deterministically depend on the time since the last subdominant interaction but these interactions are themselves stochastically distributed with a rate that depends on the ovary size of the individuals involved in them. Equation 2.39 provides a description of the system at the ovary level that retains the foremost characteristics present in the experimental data, i.e. the presence of multiple queens in the nest shortly after reprogramming and the posterior relaxation towards an analogous state to the control, as demonstrated by the results in Figure 2.16a-b. The existence of explicit time delays t_{on} and t_{off} is responsible for the transient observation of multiple queens during reprogramming. Specifically, such an overshoot arises if the value of t_{off} is of similar magnitude as the time scale associated with ovary growth.

2.8.2. Quantification of the activity stemming from fighting interactions

To quantify the activity levels of the colonies at different phases of the experiment we analyzed video recordings of individual nests (Figure 2.17a). Specifically, we quantified pixel changes between successive frames, using custom code and the computer vision library OpenCV (Figure 2.17b).

The time evolution of the global activity in a nest (Figure 2.17c-d) depends on a component which is independent of the fighting interactions involved in the regulation of the reprogramming process and on a component reflecting these interactions, being this component the only one reflected by the model. We further note that the contribution of the component depending on fighting interactions is proportional to the total interaction rate. The total interaction rate can be calculated from Equation 2.39 to be proportional to

$$\sum_{ij} o_i o_j f(o_i) f(o_j) = \left(\sum_i o_i f(o_i) \right) \left(\sum_j o_j f(o_j) \right) = \langle o_i \rangle^2.$$

Parameter	Value	Justification
ω (Interaction rate)	3.7 day^{-1}	Measurements from video recordings.
τ_{off} (Typical time between an interaction and changes in ovary growth)	2.5 days	We estimated this parameter based on the observation of 2-3 egg layers in the early commitment phase.
o_r (Ovary growth rate)	0.25 mm day^{-1}	The first egg layer is observed 6 days after queen removal with a size of mature eggs of 1.5 mm.

Table 2.1.: Summary of parameter values used for predicting experimental data.

where $\langle o_i \rangle$ represents the average ovary size across the population.

The results of our analysis are presented in Figure 2.16c. To take into account the different contributions to the global activity mentioned above, we added an offset value to the theoretical prediction such that the empirical and theoretical values matched in the control phase. Additionally, activity levels were normalized in order that the maximum of both curves matched.

2.8.3. Estimation of interaction parameters and associated timescales

The model defined in Equation 2.39 includes several parameters. In this section we provide justifications for the parameter values used to predict the experimental observables in Figure 2.16 and Figure 2.17. We have to exercise caution in interpreting these parameters in literal biological terms, as in the derivation of the model, Equation 2.39, we did not assume any non-linearities unless supported by experimental data. Such non-linearities, for example in the relation between ovary growth and gene expression, naturally exist in any biological system. We still expect, however, that the order of magnitude of parameters is not altered by unknown non-linearities and can be estimated by independent observations from experiments or the literature. The parameter values used for our predictions and their justifications are summarised in Table 2.1 and Appendix A.

2.9. Epigenetic regulation leads to an increase of the lifetime of the society by reducing gene expression variance

In previous sections we derived a continuum approach in the form of a PDE from a master equation describing a stochastic system. Although the mean-field master equation (Equation 2.22) accurately captures the salient features of the phase space of the system, it never-

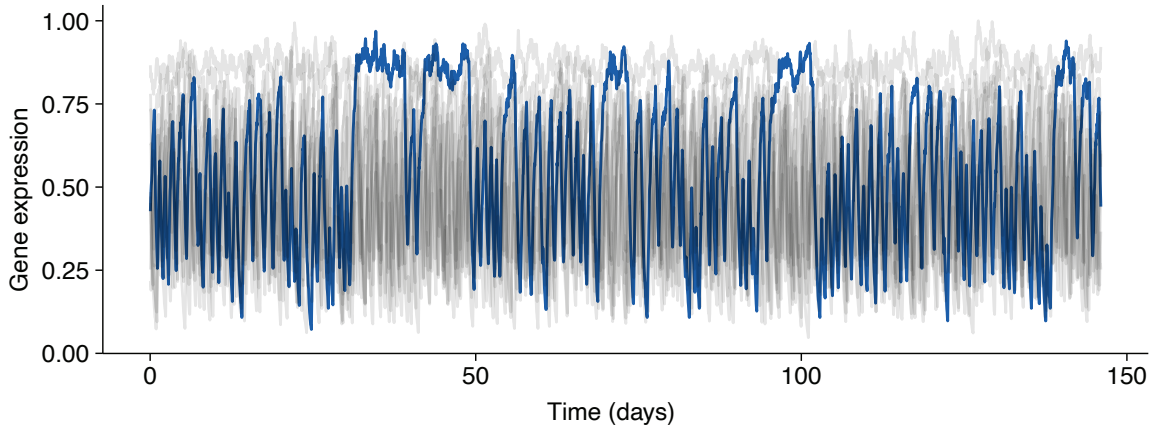


Figure 2.18.: Stochastic trajectory of the master equation Equation 2.13 shows frequent queen switching events.

theless neglects fluctuations caused by the small number of individuals present in each nest. In this section we consider how the stability of the nest and the social order can be maintained in the presence of strong fluctuations.

In Control nests, fighting interactions between the queen and workers are infrequent (Figure 2.8), happening on the same time scale as the molecular processes involved in gene regulation. The number of individuals present on a nest being small, one would expect the stochastic nature of the interactions to be a crucial factor in the dynamics of single nests. In fact, stochastic fluctuations could lead to queen switching events where the queen role changes between individuals but keeping the number of queens in the system constant. Indeed, stochastic equations of the full master equation (Equation 2.13) show frequent queen switching events (Figure 2.18).

In order to quantitatively understand how fluctuations affect the stability of the society we begin by estimating the persistence time of the society given the typical time scales of molecular and population level processes. To this end, we first calculate the probability that any individual upregulates the queen genes in the time interval between two consecutive interactions, and thereby destabilises the society, by producing a sufficient number of queen gene products. If interaction events are statistically independent, the waiting time T between consecutive interactions follows an exponential distribution, $P(T) = \omega e^{-\omega T}$. Denoting the time scale of molecular processes (e.g. the time needed for an individual to upregulate the queen genes) by T_δ , the probability that T is longer than T_δ is

$$P(T > T_\delta) = \int_{T_\delta}^{\infty} P(T) dT = e^{-\omega T_\delta}. \quad (2.40)$$

Thus, the probability of a single worker being subject to a repressive interactions in a time interval of length T_δ is $1 - e^{-\omega T_\delta}$. In a nest with N workers, the probability of all of them being subject to a repressive interaction in a time interval T_δ then is $(1 - e^{-\omega T_\delta})^N$. Conversely, the probability that at least one worker is capable of upregulating its queen genes between two consecutive interactions is

$$P = 1 - (1 - e^{-R})^N, \quad (2.41)$$

where $R = \omega T_\delta$ is a dimensionless parameter that represents the ratio between the molecular and the interaction time scales. Finally, the number of interactions before an insect is capable of upregulating its queen genes is distributed according to a geometric distribution with mean $1/P$. Therefore, the persistence time of the society τ is

$$\tau = \omega^{-1} (P^{-1} + R). \quad (2.42)$$

The theoretical prediction of Equation 2.42 indicates that in order to obtain a queen turnover time consistent with the experimentally measured value of roughly 27.5 weeks [81] the ratio between molecular and interaction time scales, R , must be larger or equal than 9 (Figure 2.20a). Given the experimental estimation of the time interval between two consecutive interactions, this would imply that gene expression states are stable for at least three days in the absence of interactions. However, transcriptional queen signatures are already established in *Polistes* three days after queen removal [65].

Furthermore, experimentally it is observed that the queen phenotype is stable over the whole lifetime of the organism, as the only natural queen replacement mechanisms are queen death or removal. Therefore, a mechanism that reduces fluctuations is required in order to stabilize phenotypes over long periods of time.

Organisms have developed mechanisms that regulate gene expression without modifying the nucleotides of the DNA sequence. These mechanisms and their study are known as epigenetics [82]. In particular, the primary layer of epigenetic modifications is DNA methylation [82, 83]. DNA methylation is a chemical modification of the DNA that predominantly affects cytosines located next to a guanine (CpG) in the DNA sequence. When a cytosine is methylated a methyl group is attached to it, altering the chemical activity of the corresponding DNA fragment without modifying the sequence itself. DNA methylation can be experimentally measured genome-wide using whole-genome Bisulfite sequencing [84], a technology that modifies methylated cytosines leaving unmethylated bases unaltered, allowing for the detection of methylation marks.

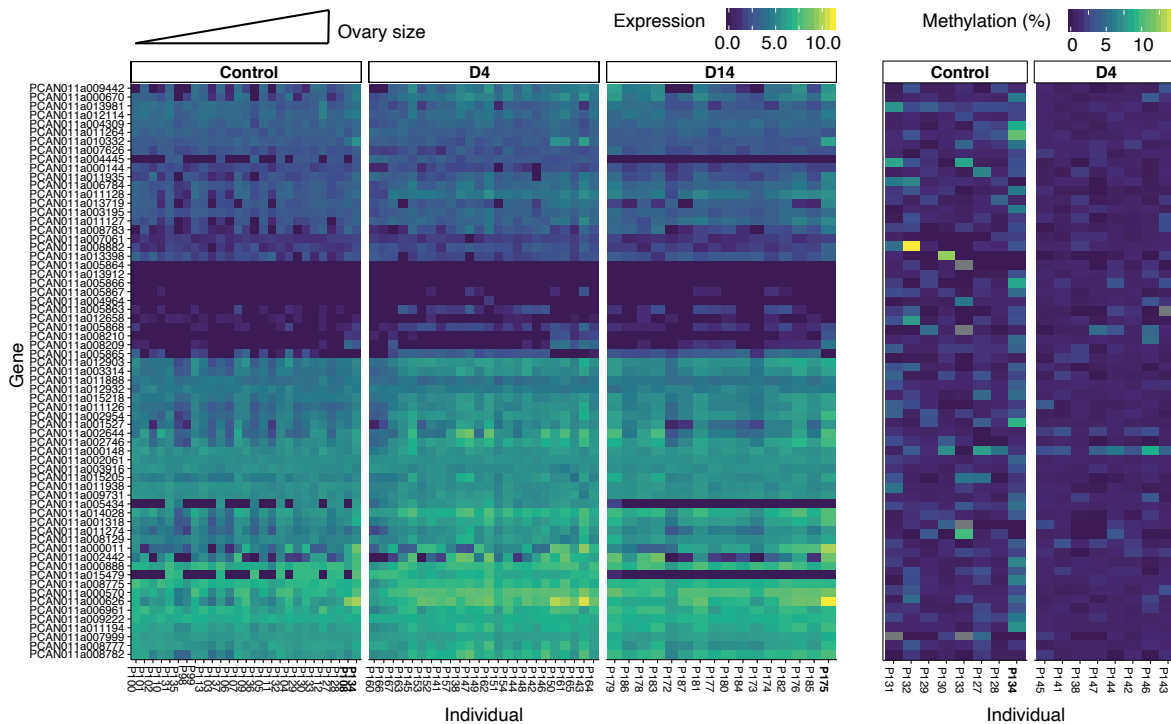


Figure 2.19.: Heatmaps of gene expression (left) and DNA methylation (right) levels of queen genes. Genes are ordered based on hierarchical clustering of gene expression in control nests. Individuals are ordered by their ovary size (small to large, left to right) and queens are marked in bold.

In mammals, DNA methylation plays a crucial role during development [85–87], as the removal of methylation marks of the parental genome is followed by the imprint of new marks during the initial developmental stages, a process that has to be tightly regulated as errors in the imprinting of new methylation marks have been associated with the appearance of cancerous cells [88, 89]. Additionally, DNA methylation has been shown to influence gene expression, including silencing undesired genes at different developmental stages [90, 91].

In recent years, the existence and the role of epigenetic mechanisms of genomic regulation have been the focus of studies in social insects [92]. In particular, gene bodies are susceptible to being methylated in *Polistes canadensis*, as reflected by the measured methylation levels of the queen genes (Figure 2.19).

To understand how DNA methylation can suppress fluctuations that destabilise *Polistes* societies we first studied the relationship between expression and methylation of all genes in *Polistes* (Figure 2.20b). Our results reveal that, besides the existence of a positive correlation between gene expression and methylation level, there is a correlation between gene variability across individuals and the levels of DNA methylation of those genes (Figure 2.21a), with genes that are methylated being less variable than those which are not. Motivated by this evidence, we hypothesized that the regulation of gene expression by methylation marks could act as

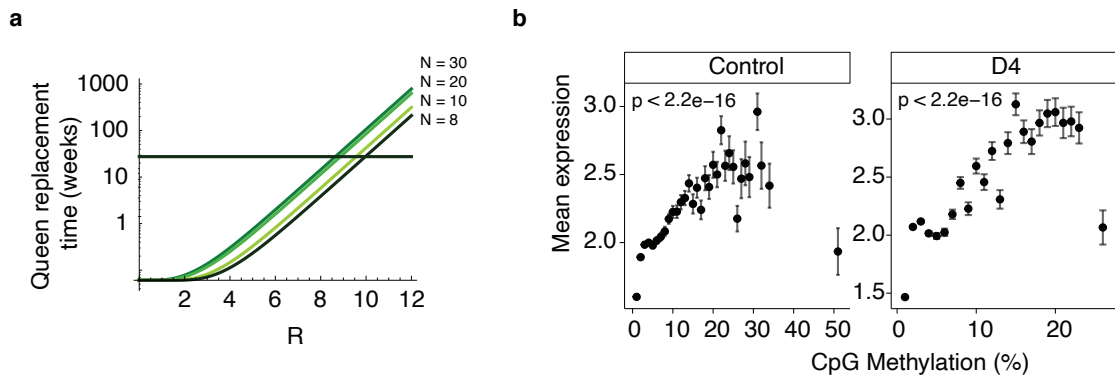


Figure 2.20.: **a)** Predicted queen replacement times as a function of the ratio R between interaction and molecular time scales for different values of the population size N . The horizontal line indicates the experimentally measured queen lifespan in *Polistes canadensis*. **b)** Average normalized expression as a function of the percentage of methylated CpGs. Dots correspond to the average over all genes with the same DNA methylation percentage. p -values were determined from a Pearson correlation test.

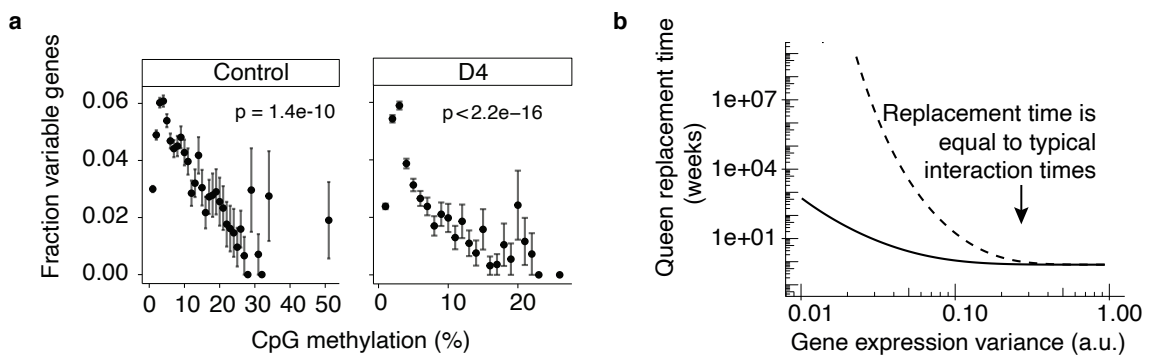


Figure 2.21.: **a)** Fraction of variable genes as a function of the percentage of methylated CpGs. **b)** Theoretical prediction of queen replacements times as a function of gene expression variance across insects for exponentially (solid line) and normally (dashed line) distributed queen gene expression levels.

a stabilizing mechanism that increases the queen lifetime by decreasing transcriptional noise and reducing gene expression variability across the population.

To test this hypothesis, we begin by considering the steady-state probability distribution of queen gene expression values, $P(r)$, in workers. In a population composed of N workers, the probability of observing a new queen is equal to the probability that at least one worker upregulates the queen genes. Assuming that the queen gene expression values in workers are distributed according to $P(r)$, the probability of observing a new queen is equal to the probability of finding an insect with queen gene expression larger or equal to one,

$$p = \int_1^{\infty} P(r) dr. \quad (2.43)$$

As per our previous calculation, Equation 2.42, the persistence time of the society then is

$$\tau(\sigma) = \omega^{-1} \left[\frac{1}{1 - (1 - p)^N} + R \right]. \quad (2.44)$$

The variance of $P(r)$, σ^2 , is determined by an interplay between queen gene expression and queen interactions. Our main conclusions from this analysis are independent of the details of these processes. We here provide two examples to illustrate our results. If subdominant interactions with the queen are statistically independent and happen at a rate independent of queen gene expression levels in workers, then queen gene expression levels in workers follow an exponential distribution, $P(r) = \exp(-\omega r/\mu)$, with μ and ω being the rate of queen gene expression and interactions, respectively. With the variance given by $\sigma^2 = (\mu/\omega)^2$ we obtain $p = \exp(-1/\sigma)$ and a persistence time of the society of

$$\tau = \omega^{-1} \left[\frac{1}{1 - (1 - e^{-1/\sigma})^N} + R \right]. \quad (2.45)$$

In a more realistic setting corresponding to our model, the rate of repressive interactions of a worker is proportional to its level of queen gene expression. In this case, in analogy to stochastic growth processes with resetting, queen gene expression levels of workers follow a truncated normal distribution [93],

$$P(r) = \sqrt{\frac{2\omega^2}{\pi\mu^2}} e^{-\frac{r^2\omega^2}{2\mu^2}}, \quad (2.46)$$

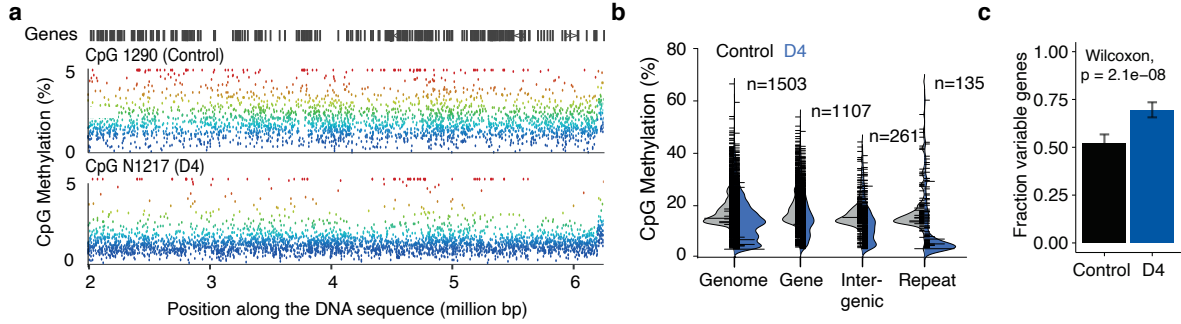


Figure 2.22.: **a)** Example of global demethylation. Every dot represents average methylation across individuals in a genomic window containing 50 informative CpGs. Colour reflects methylation level of the probe. **b)** Average of CpG methylation levels for different genomic features across 8 individuals collected before (eggless phase) and after (D4 phase) queen removal. Only windows containing 50 informative CpG with at least 10 % average methylation in at least one phase are shown. **c)** Fraction of significantly variable genes (p -value < 0.1 , t -test) in Control and D4 phases.

with variance $\sigma^2 = \mu^2/\omega^2$. In this case we find that

$$\tau = \omega^{-1} \left(\frac{1}{1 - (1 - \text{Erfc}[(2\sigma^2)^{-1/2}])^N} + R \right), \quad (2.47)$$

where $\text{Erfc}(x) = 1 - \text{Erf}(x)$ is the complementary error function.

Our results (Figure 2.21b) indicate that a decrease of gene expression variability across the population leads to an exponential increase of the lifetime of the society. Therefore, the imprinting of methylation marks in the individual genomes can have a strong stabilising effect at the population level, indicating a possible novel role for epigenetics mechanisms in the formation and maintenance of primitive societies.

This conclusion is further supported by the demethylation of all genomic regions (Figure 2.22a-b) after queen removal, accompanied by an increase of the fraction of variable genes (Figure 2.22c) during the reprogramming process.

A corollary of our theory is that the stabilization of larger societies would require a stronger decrease in the inter-individual variance, corresponding to a larger average methylation level of the genome. In order to test this prediction we did an inter-species comparison between *Polistes canadensis*, *Belonogaster juncea*, a quasi-social species of wasps, as well as the data from [92] (Figure 2.23). A quasi-social insect, *Belonogaster* wasps form small colonies where different individuals exhibit the queen phenotype in a single generation [94]. Accordingly, we found an overall lower methylation level in *Belonogaster* wasps compared to *Polistes canadensis* (Figure 2.23a). This is corroborated by a positive correlation (Figure 2.23b) between global levels of DNA methylation and queen replacement times and population sizes across several

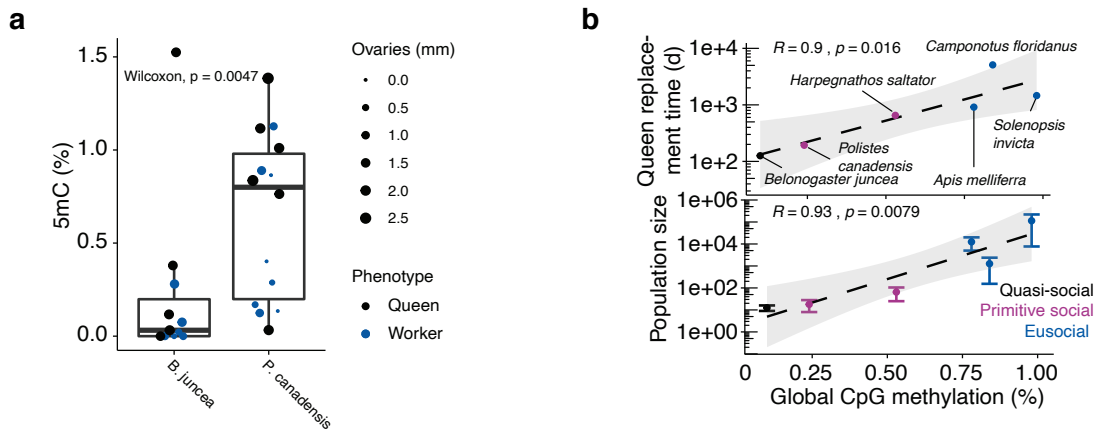


Figure 2.23.: **a)** Global level of DNA methylation measured by mass spectrometry in *Polistes canadensis* and *Belonogaster juncea* control nests. Center line corresponds to the median and lower and upper hinges to the 25th and 75th percentiles, respectively. Only nests with at least 2 individuals collected are shown. **b)** Top: Queen replacement times for different social insect species as a function of the DNA methylation percentage of the genome. Bottom: Population size for different social insect species as a function of the DNA methylation percentage of the genome. Color indicates the degree of sociality.

species of social insects. These results indicate that more complex social structures regulated by stochastic interactions would require additional regulation from epigenetic processes in order to be stable over long periods of time.

2.10. Conclusions

Through the combination of experimental and theoretical approaches we have shown that *Polistes* uses antagonistic dynamics on different spatial scales to distinguish between intrinsic and extrinsic perturbations, being stable against the former and reacting plastically to the latter. Owing to the unique experimental strategy (Figure 2.2), we could correlate observations ranging from the molecular to the population scales in order to build a theoretical model that captured the most relevant features of the biological system and set out to study its response to perturbations. Our results uncover how the remodelling of microscopic degrees of freedom by macroscopic feedback gives rise to specialisation and plasticity in this primitive society.

To study the effects of different perturbations to the society, colonies of *Polistes canadensis* were studied in their natural habitat during field work expeditions in Panama. While studying a social insect in its natural environment limits the use of molecular perturbation techniques available in laboratories, it allowed us to perform behavioural perturbations under natural

environmental fluctuations of *Polistes* nests, preserving their contribution to the dynamics of the system.

An accurate theoretical description of the biological complexity of the processes regulating the insect society would not allow gaining analytical insight into the response of the society to perturbations. We therefore based our theoretical model on the minimal set of assumptions that followed directly from the experimental observations. Our theoretical work shows that specialisation and plasticity in *Polistes* are emergent properties arising from the active feedback between microscopic and macroscopic degrees of freedom. In particular, the possible phenotypes of the system correspond to the fixed points of a dynamical system that undergoes a bifurcation where the role of control parameter is played by the population structure (Figure 2.12). As a consequence, after a perturbation of the population structure the balance between the molecular and population terms of the dynamics is temporarily broken, resulting in the observed reprogramming dynamics. These analytical insights are corroborated by the results of stochastic simulations of the full non-Markovian dynamics, that additionally provide quantitative predictions of experimental observables. Our results show how societies of primitively social insects can control how fluctuations propagate across scales of biological organisation to perform specific functions.

Furthermore, we explored the effect of stochastic fluctuations in the system dynamics, revealing that stable societies over long periods of time must rely on additional regulatory process. Our results suggest that DNA methylation might play an unanticipated role in regulating the stability of the society at the colony level (Figure 2.21). Correlating DNA methylation measurements with gene expression, we discovered that methylated genes are less variable than unmethylated genes. We predicted that this reduction in variability could lead to an exponential increase of the lifetime of the society, introducing the possibility that epigenetic modifications of individual genomes have a relevant effect at the population level.

In conclusion, our work demonstrates that correlated measurements across scales can give qualitatively new insights into the mechanisms underlying self-organisation of biological systems. The approach developed here may be more widely applicable to other biological systems of interest, leveraging the possibilities offered by the advent of new experimental and computational techniques. In Chapter 4 we discuss in more detail future research lines motivated by our work on this fascinating insect society.

3. Driving stochastic lattice systems by reinforcement learning

What is love except another name for the use of positive reinforcement?

B. F. Skinner - Walden Two

Social insects have provided a fertile ground to study the remodelling of microscopic degrees of freedom by macroscopic feedback in biological systems. In Chapter 2 we have shown that primitively social paper wasps employ antagonistic processes at the molecular and the population scales to control the propagation of fluctuations across spatial scales. The control of fluctuations and the response of the system against perturbations is a central research question in several fields, not only in biology. For instance, in the field of artificial intelligence it is desired that systems are robust against noise in the input signal and perturbations of the system structure.

In this chapter we explore the interplay between microscopic and macroscopic degrees of freedom in artificial intelligence systems. In particular, we define a many-particle stochastic model, where the microscopic degrees of freedom of individual particles are remodelled as a result of macroscopic feedback. Making use of tools of the field of deep learning, we represent the microscopic degrees of freedom of single particles using deep neural networks that are dynamically remodelled following stochastic learning dynamics. In this section we provide a historical overview of reinforcement learning and motivate how the remodelling of microscopic degrees of freedom by macroscopic feedback in artificial intelligence systems can be studied through the combination of macroscopic stochastic dynamics and neural networks trained using reinforcement learning.

One extraordinary feature of biological systems is the ability to detect environmental changes and react to them accordingly, as exemplified by the evolution of species over long time scales and by learning over the lifetime of a single individual.

Since long before the origins of scientific thought, learning, understood as the acquisition of new behavioural patterns and knowledge over the lifetime of an individual, and the processes governing it have drawn the attention of a great number of scholars, from Aristotle to Francis

Bacon [95, 96]. Later on, with the advent of modern science, learning processes became one of the central objects of studies of fields like psychology and neuroscience, focused respectively at the behavioural and physiological aspects of these processes.

In particular, one of the most widely accepted learning paradigms is offered by behaviourism [97, 98], a psychological school of thought that understands learning as a process where behavioural patterns are reinforced or extinguished based on the response to those behaviours, supported by strong experimental evidence [99, 100].

The power and influence of the ideas of behaviorism introduced above went beyond the field of psychology, deeply impacting the development of theories of artificial learning, such as statistical learning theory and machine learning (ML) in the decades of the 50s and 60s [101]. In particular, a subfield of ML, the so called reinforcement learning (RL), draws direct inspiration the ideas of behavioural psychology. In short, a RL algorithm iteratively learns probability distributions over a certain action space that maximize a given reward function. The reward function depends on the state of the system and the action taken. After every time step, the state of the system is updated and the reward is computed. As a result of the training dynamics, in analogy with behavioural psychology, actions leading to a higher reward are reinforced whereas those leading to lower rewards are extinguished. The appeal of RL compared to other disciplines of ML relies on the wealth of behavioural patterns that can result from the maximization of simple reward functions [102].

An active field of research since its inception, the interest in RL has nevertheless sharply increased in the last decade. The application of deep artificial neural networks to solve RL tasks [103], also known as deep reinforcement learning, has led to an increase of performance in a wide array of classical RL tasks, including game playing [52, 103, 104] or training autonomous robots [105].

Deep reinforcement learning has not only revolutionized theoretical computer science. In 2018, AlphaZero [53], a game-playing engine powered by deep neural networks, made headlines around the world as it was able to soundly defeat human players as well as the best engines to date (Figure 3.1) in three board games as different as Chess, Shogi and Go. Furthermore, AlphaZero learned to play the three games from self-play and did not rely in any heuristic rules. A remarkable feat on its own, AlphaZero relied on the same network architecture to play all three different games, therefore being much less specialized than the engines it competed against. In the world of chess, the impact of AlphaZero and its aggressive play has led the development of new opening ideas and long term strategic plans [106]. Top-level players and amateurs alike now include the assistance of neural network based chess engines

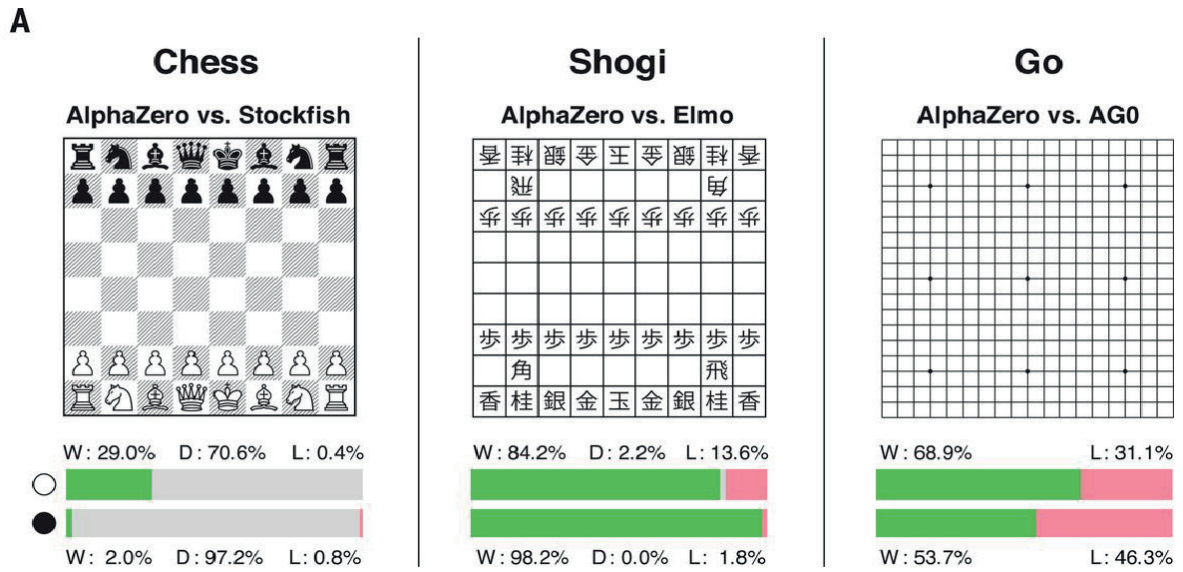


Figure 3.1.: Performance of AlphaZero in terms of win, draw and loss percentages against the top engines in **a)** Chess, **b)** Shogi and **c)** Go. Adapted from [53]

in their training regimes, as they often provide different evaluations from those of traditional engines based on heuristic rules, offering the player fresh, unexplored ways to play.

The recent introduction of deep learning techniques to study physical systems [42, 43, 107, 108] prompts the question of whether such advances can have a similar impact in physics. The statistical physics community has devoted serious efforts to derive mechanistic interpretations of deep learning [109–111] and to characterize the dynamics of these systems [112–114]. Furthermore, in recent years the intimate connection between RL and optimal control problems has also been leveraged to gain insight into this class of systems [115–118].

On the other hand, the intersection between deep learning and reinforcement learning has not received the same attention from the physics community, although the successful application of deep reinforcement learning to statistical physics problems defines this as a promising line of research. In particular, deep reinforcement learning has been successfully applied to a long studied non-equilibrium physics problem, the prediction of protein folding from the sequence data, with the performance of the aptly named AlphaFold [119–121] above any other available algorithm or theory.

In this chapter, we employ tools from deep learning to define a stochastic many-particle model where we characterise how microscopic degrees of freedom are remodelled as a result of macroscopic feedback. In particular, we consider a system of N particles, each with several associated intrinsic degrees of freedom (Figure 3.3), that we represent using deep neural networks. The dynamics are defined by the reaction rates of individual particles, determined by the output of the associated neural networks. Therefore, within this framework,

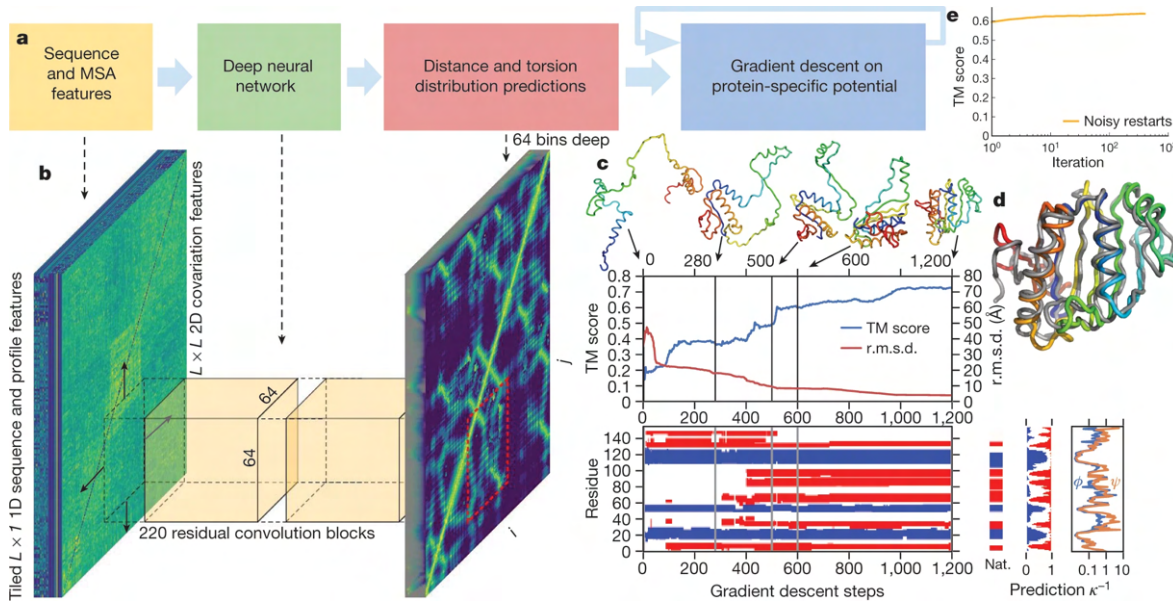


Figure 3.2.: Illustration of the AlphaFold prediction process for a specific protein. Adapted by permission from Springer Nature: Nature (Improved protein structure prediction using potentials from deep learning, Senior, A.W. *et al.*), (2020) [119]

the microscopic variables correspond to the parameters of individual neural networks and the macroscopic level is characterized by the particle dynamics. In order to couple the microscopic and the macroscopic scales, the input of the neural networks is given by the state of a neighbourhood of the corresponding particle and the neural network parameters are updated using reinforcement learning based on the trajectory of the system.

Reinforcement learning provides a suitable learning paradigm to induce couplings between particles resulting from the existence of a shared environment, as the dynamics of a single particle modify the state of the neighbourhood of other particles. As a result, the reaction rates of single particles depend on the dynamics of all other particles in the system and are updated depending on the history of the dynamics. Specifically, Figure 3.4 schematically illustrates the coupling between deep reinforcement learning and stochastic many-particle dynamics in our system. The effective feedback between scales resulting from the interplay between the macroscopic and the learning dynamics, makes this a well-suited system to characterise how microscopic degrees of freedom are remodelled by macroscopic feedback.

This chapter is structured as follows: first, in Section 3.1 we introduce the required ML concepts, focusing on RL, deep learning and deep reinforcement learning. Then, we introduce our model and the specific details of its numerical implementation in Section 3.2. In the following sections we quantify the resulting dynamics at the macroscopic (Section 3.3, Section 3.4) and microscopic (Section 3.5) levels, corresponding to the trajectories of individual particles and collective properties of the system and the neural networks structure,

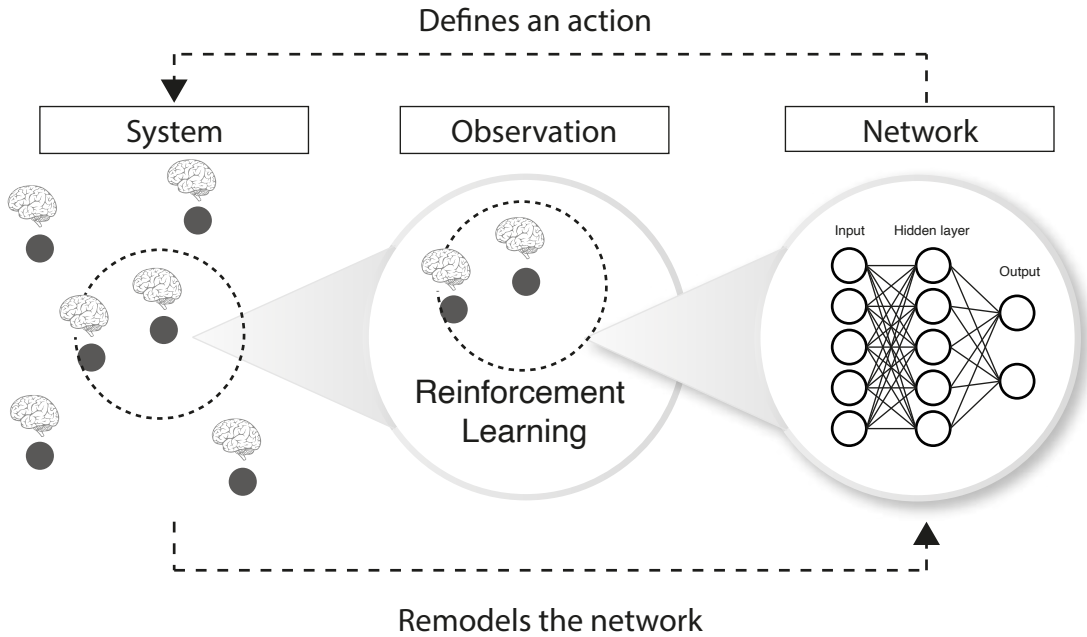


Figure 3.3.: Schematic description of the feedback between the macroscopic and microscopic degrees of freedom in the system.

respectively. In Section 3.6 we characterise how the interplay between the microscopic and the macroscopic dynamics gives rise to the emergence of complex interactions over time. Later, we study how the dynamics are impacted by the introduction of memory (Section 3.7) and disorder (Section 3.8). Finally, in Section 3.9 we conclude with some summarizing remarks.

3.1. An introduction to deep reinforcement learning

In this section we introduce the framework of deep reinforcement learning as the combination of reinforcement learning and deep learning, focusing on its mathematical foundations and computational implementations. In addition, we detail the specific scheme that we later implement in our numerical simulations.

3.1.1. Reinforcement learning

Reinforcement learning [51] refers to a class of machine learning methods where an agent is trained to find the optimal policy (sequence of actions) that maximizes a reward function. In layman terms, each action is associated with a reward and the algorithm is trained to find the actions that result in the maximum return. Therefore, to characterize a RL task we need to define an agent, the action set, the reward function and the environment the agent acts upon.

The dynamics of such a task are defined in discrete time steps as follows: at each step the

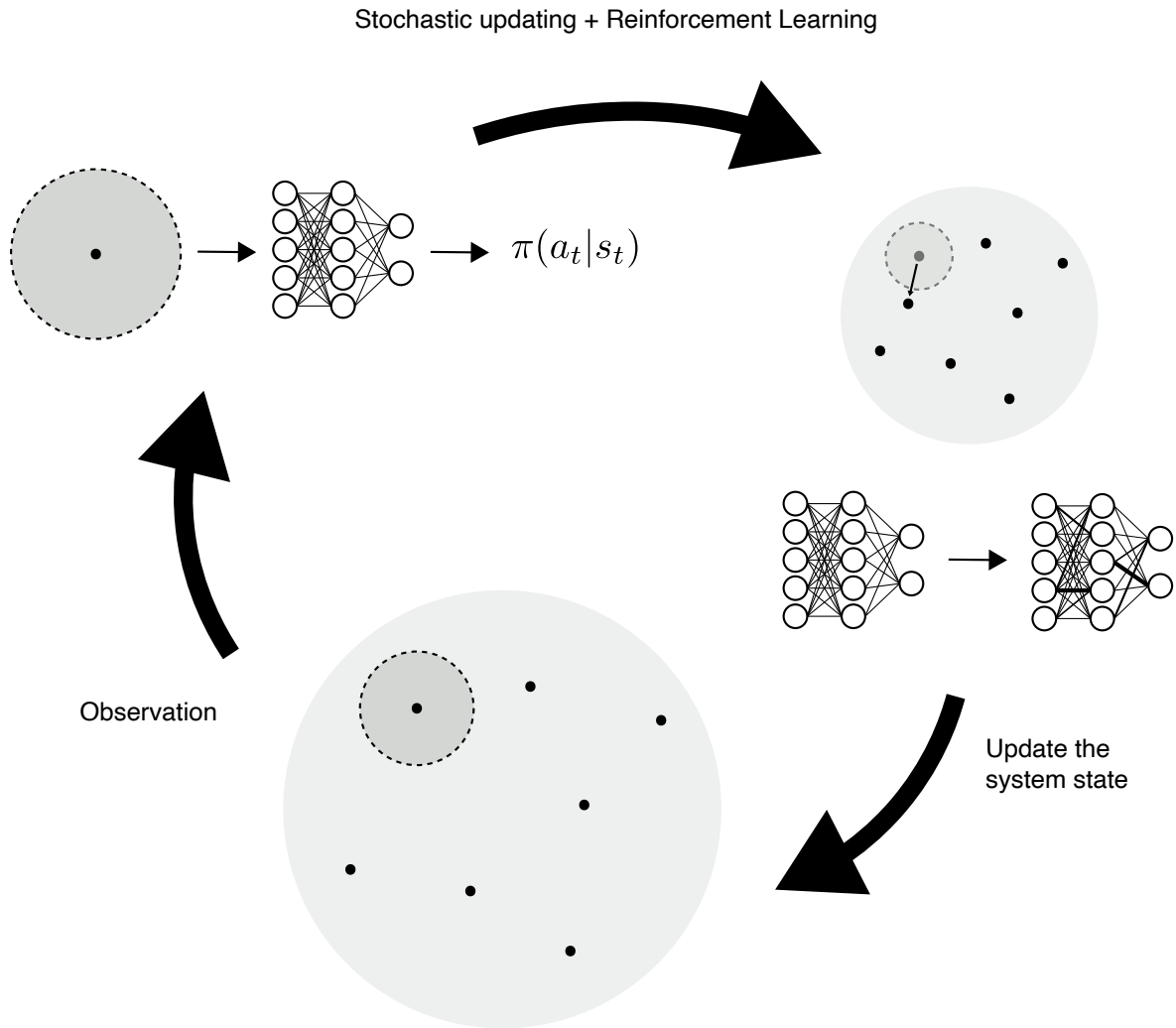


Figure 3.4.: Schematic depiction of the coupling between stochastic and learning dynamics in the system.

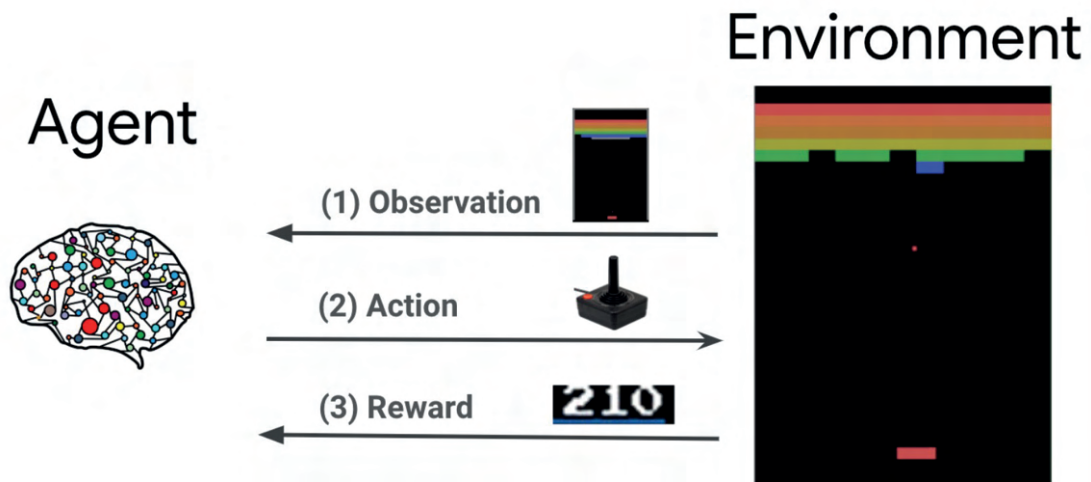


Figure 3.5.: Elements of a reinforcement learning task. Adapted from https://www.tensorflow.org/agents/tutorials/0_intro_rl.

agent performs an action on the environment based on the previously received observation from the environment and, as a result, receives a reward and the next observation from the environment. In the following we mathematically define the policy, value and cumulative reward functions and discuss how the optimization of the policy is typically achieved.

The policy function, denoted as π , determines the action taken by an agent when presented with a particular observation of the environment. Policies can be either deterministic, where the action $a_t = \pi(s_t)$ is unequivocally determined given the current environmental state s_t , or stochastic, describing the conditional probability of the agent selecting a particular action given the state of the environment $\pi(a_t|s_t)$. By randomly selecting the next action, stochastic policies are less likely to get stuck in local minima of the loss landscape, what is known in the reinforcement learning literature as the exploitation-exploration tradeoff [51]. In the following we consider stochastic policies, as deterministic policies are a limit case of these.

In order to quantify and compare the performance of different policies, we define the state-value function associated with the policy π as

$$V_\pi(s) = \mathbb{E}[R|s_t = s], \quad (3.1)$$

where R is the expected return that an agent would get if the system is in state s and the agent follows policy π . The optimal policy π^* is defined as the policy that maximizes the expected return

$$\pi^* = \operatorname{argmax}_\pi V_\pi(s). \quad (3.2)$$

Although the state-value function is enough to define the optimal policy, in practice it is also useful to define the state-action value function Q_π

$$Q_\pi(s, a) = \mathbb{E}[R|s_t = s, a_t = a], \quad (3.3)$$

representing the expected return of an agent that first executes action a and then follows the policy π . In several applications this allows for computing the optimal policy in an easier way, taking advantage of Bellman's optimality equation [122] to update the state-action value function.

To complete the above definitions we need to define the return R associated to a given policy. The return R is defined as the cumulative sum of discounted future rewards

$$R = \sum_{t=0}^T \gamma^t r_t, \quad (3.4)$$

where r_t is the reward obtained at time t and $\gamma \in [0, 1]$ is the discount parameter. The discount parameter measures the weight assigned to immediate compared to future rewards in the updating of the policy. For $\gamma = 0$ only instantaneous rewards are relevant for the updating of the policy whereas for $\gamma = 1$ all future rewards are weighted equally. In our deep RL framework the sum in Equation 3.4 is taken over the length of an episode, corresponding to the time interval between neural network updates. More details about our specific implementation are given in Section 3.2.

Once all elements of a RL task have been defined, all that remains is optimizing the policy function π . The large dimensionality of the action and state spaces in practical applications renders analytical calculations of the optimal policy unfeasible. Therefore, one of the main focuses of classical RL has been the development of algorithms that approximate the optimal policy [123–126]. Nevertheless, the lack of generalization and scalability of classical methods [51] makes them not suited for complex applications with large amounts of data. In connection to the recent advances in the application of artificial deep neural networks as function approximators, in recent years deep reinforcement learning [52, 103] has emerged as a promising way of approaching RL tasks. The next section is devoted to introducing deep learning and discussing its application to RL problems.

3.1.2. Deep learning and its application to reinforcement learning tasks

Deep learning [41] is a sub-field of machine learning characterized by the use of deep artificial neural networks whose methods have been successfully applied in a broad range of problems, including speech recognition, machine translation, bioinformatics or structural protein sequence analysis [127–132]. In this section we introduce the main component of deep learning, deep neural networks, their role as function approximators and their use in RL tasks.

From a mathematical perspective, deep learning comprises a series of methods capable of learning probability distributions from data [43]. Deep learning methods are traditionally divided in three different categories (supervised, semi-supervised and unsupervised) depending on the probability distribution being learned. In particular, given data x , unsupervised learning methods are concerned with approximating the probability distribution underlying the data $p(x)$, whereas supervised learning methods approximate a conditional probability distribution $p(x|y)$ given the data x and labels y .

The central component of any deep learning algorithm are deep (artificial) neural networks. A deep neural network is an artificial neural network with hidden layers between the input and the output layers (Figure 3.6). Depending on the connections between nodes and layers,

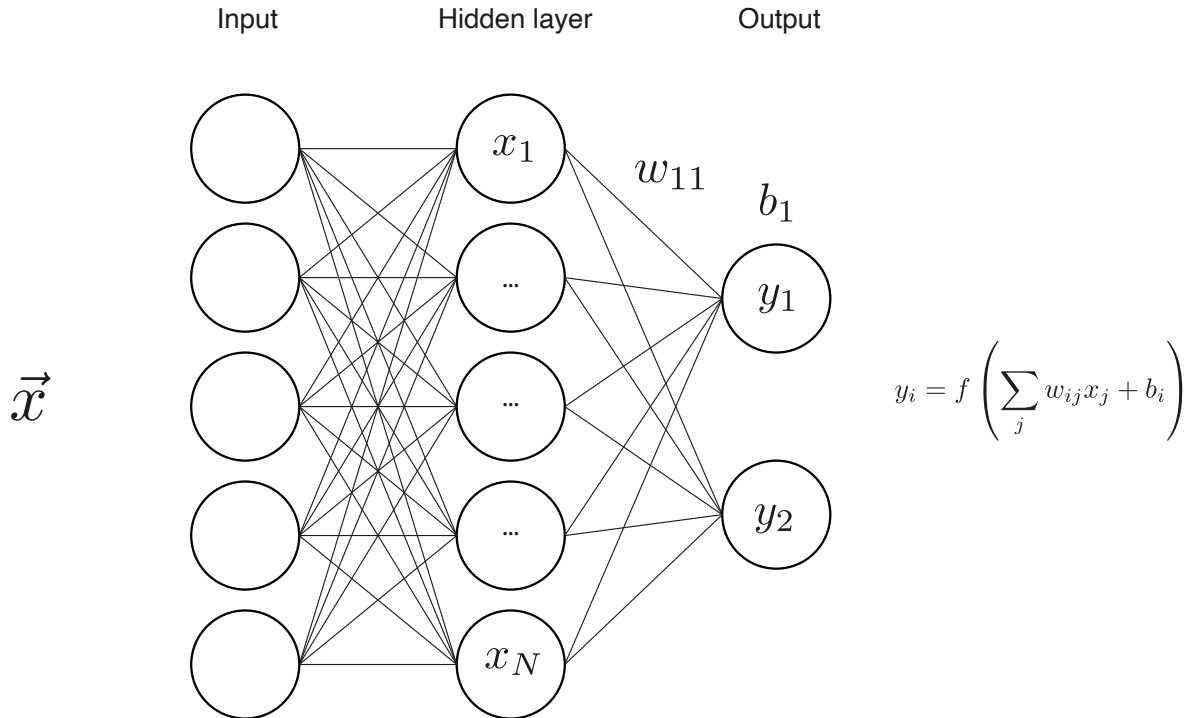


Figure 3.6.: Structure of a feed-forward neural network. Labels illustrate the calculation of the output for units in the output layer.

and in order to exploit the symmetries of the task at hand, several types of artificial neural networks have been defined, including convolutional neural networks (CNN) [133], recurrent neural networks (RNN) [134] or feed-forward neural networks. The power of deep learning and deep neural networks is based on the universal approximator property of deep neural networks [135–137], referring to the fact that any arbitrary function can be approximated using a suitably chosen deep neural network with non-linear activation functions.

In the following we consider only feed-forward neural networks, as this architecture will correspond to the main building blocks of our model. This means that we consider networks composed of an input layer, a finite number of hidden units and a final output layer where the units of a given layer are connected only to the previous and following layers (Figure 3.6). Feed-forward neural networks are characterized by their parameters, weights and biases $\theta = (\mathbf{w}, \vec{b})$, and a non-linear activation function f that together determine the output of a unit as a non-linear function of a linear combination of the inputs coming from the previous layer,

$$y_i = f\left(\sum_j w_{ij}x_j + b_i\right), \quad (3.5)$$

where the sum is performed over all units in the previous layer, w_{ij} is the weight that unit i assigns to the input from unit j and b_i corresponds to the bias of unit i . Common choices

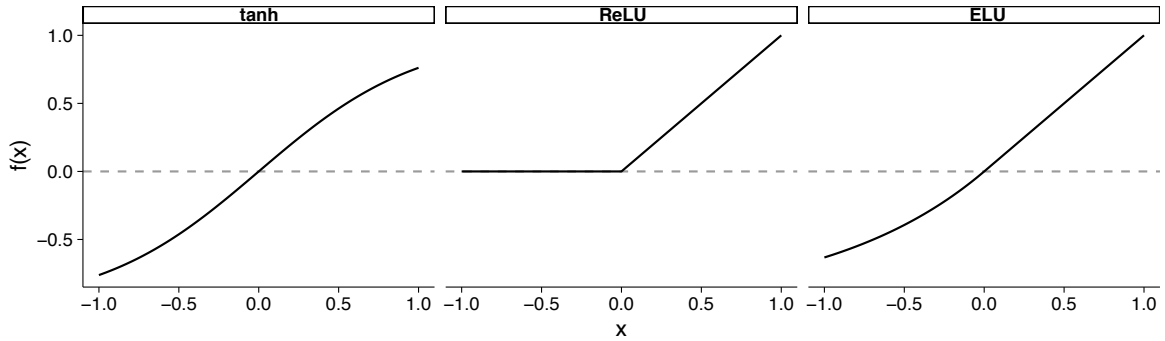


Figure 3.7.: Shapes of common non-linear activation functions: **a)** Tanh. **b)** Rectified linear unit (ReLU). **c)** Exponential linear unit.

for the non-linear activation function are sigmoidal, rectified linear units (ReLU) or tanh functions (Figure 3.7).

During training the parameters of the neural network are updated to minimize a suitably chosen loss function, $\mathcal{L}(\theta)$. We will discuss the choice of the loss function in Section 3.2, where we define our reinforcement learning driven lattice gas. Next, we consider the optimization of the loss function with respect to the parameters θ . One of the most commonly used optimization methods is gradient descent, corresponding to updating the parameters according to the negative gradient of the loss function, eventually reaching a local minimum of the loss landscape. The simplest gradient descent update rule for the parameters can be written as

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}, \quad (3.6)$$

where $\nabla_{\theta} \mathcal{L}$ is the gradient of the loss function with respect to the parameters θ and η is the learning rate and controls the step size in the direction of the gradient.

Nevertheless, the use of classical gradient descent to train neural networks has a number of shortcomings. As the size of the datasets and the number of parameters commonly used in deep learning applications grows, computing the gradient of the loss landscape with respect to the parameters becomes computationally very expensive [43]. Furthermore, deterministic optimization algorithms like gradient descent can get stuck in a local minima of the loss landscape depending on the initial conditions. In order to mitigate these shortcomings, neural networks are trained using a combination of stochastic dynamics and backpropagation to compute the gradient of the loss with respect to the parameters that we refer to as stochastic gradient descent (SGD) [138, 139].

One of the most common ways to include stochasticity in the dynamics is stochastic gradient descent with mini-batches, the method that we will implement in our numerical simulations.

The idea behind stochastic gradient descent with mini-batches is to divide the dataset into smaller subsets, referred to hereafter as mini-batches, and approximate the calculation of the gradient by using only the mini-batch to compute the loss landscape. Formally, the gradient descent equation (Equation 3.6) becomes

$$\theta_{t+1} = \theta_t + \eta \nabla_{\theta} \mathcal{L} \approx \theta_t + \eta \nabla_{\theta} \mathcal{L}_i, \quad (3.7)$$

where \mathcal{L}_i is the minibatch loss function. Approximating the gradient of the loss function using mini-batches can greatly speed up the calculations for large datasets, as in every training step only a small subset of the dataset is used.

The continued search for improved performance and lower computational cost has led to several versions of the original SGD with mini-batches method being proposed in the literature, including second order and/or momentum terms. In particular, to train our neural networks we used a second order method that includes inertial terms (ADAM) [140]. Explicitly, the parameter update rule following a step of the ADAM algorithm is given by

$$m_t = \beta_1 m_{t-1} - (1 - \beta_1) \nabla_{\theta} \mathcal{L}, \quad (3.8)$$

$$s_t = \beta_2 s_{t-1} - (1 - \beta_2) (\nabla_{\theta} \mathcal{L})^2, \quad (3.9)$$

$$\theta_{t+1} = \theta_t - \eta_t \frac{\hat{m}_t}{\sqrt{\hat{s}_t + \epsilon}}, \quad (3.10)$$

where m_t and s_t are the estimates of the first and second moments of the gradient, β_1 and β_2 the memory lifetime of the first and second moments, $\hat{m}_t = m_t / (1 - \beta_1)$, $\hat{s}_t = s_t / (1 - \beta_2)$ their unbiased corrections, η_t is the time-dependent learning rate and ϵ is a small regularization constant introduced for the sake of numerical stability.

Finally, after having introduced the central elements of a deep learning algorithm, in the following we discuss the application of deep learning tools to reinforcement learning tasks, known as deep reinforcement learning. Deep reinforcement learning is based on the idea of using deep neural networks to approximate elements of a reinforcement learning task, such as the policy function $\pi(a_t | s_t)$.

In particular, we decided to implement a numerical scheme that not only approximates the policy but also the value function. We implemented an asynchronous actor-critic scheme (A3C) [104] where two deep neural networks, referred to as actor and critic, are used to estimate the policy and the value functions. The actor decides the actions of the agent while the critic evaluates these decisions by approximating the value function. The information from the two networks is then used to update and improve the predictions of the other network.

In Section 3.2.1 we discuss the details of this deep reinforcement learning algorithm.

We refer the interested reader to [41] for an in-depth introduction to machine learning and deep learning in particular. For an introduction to reinforcement learning we refer to [51]. Finally, approaches to ML from the point of view of statistical physics and applications of ML to physical systems are discussed in [42, 43].

3.2. Deep reinforcement learning driven lattice gas

In the previous section we have introduced the required deep learning techniques to build a many-particle non-equilibrium system capable of learning from its previous history. The goal of this section is to build a minimal stochastic many-particle system where transition rates are determined by deep neural networks trained using reinforcement learning. For the sake of clarity, in the following we will focus on the case of Markovian dynamics.

We consider a system of N particles undergoing stochastic dynamics defined by the individual transition rates that are determined, at every step, by a deep neural network associated to the corresponding particle. The input of neural network i is given by the occupancy of a neighbourhood of the position of the associated particle $\mathcal{B}_r(x_i)$, where r is the radius of the neighbourhood and x_i the position of particle i . On the other hand, the output corresponds to the probability of the different actions that the particle can take. The neural networks are then updated using deep reinforcement learning to maximize the cumulative return. As the input of neural network i depends on the position of all other $N - 1$ particles, particles are coupled as a result of a shared environment. A schematic representation of the model can be found in Figure 3.8.

A minimal implementation of this paradigm is given by a one-dimensional system of N particles in a lattice of size L with periodic boundary conditions. In this case, the action space of each particle is given by \mathbb{Z}_2 , with the two elements of this field corresponding to the two directions a particle can take.

The learning dynamics are implemented as individual actor-critic architectures (Section 3.2.1) for each particle. The actor networks determine the reaction rates (probability of going left or right) of each particle, effectively coupling the macroscopic and the learning dynamics. We do not consider any additional interactions between particles, meaning that any collective dynamics are an emergent property of the interactions induced by the reward function and the shared environment.

For a given state of the actor networks, the macroscopic dynamics of the system can be

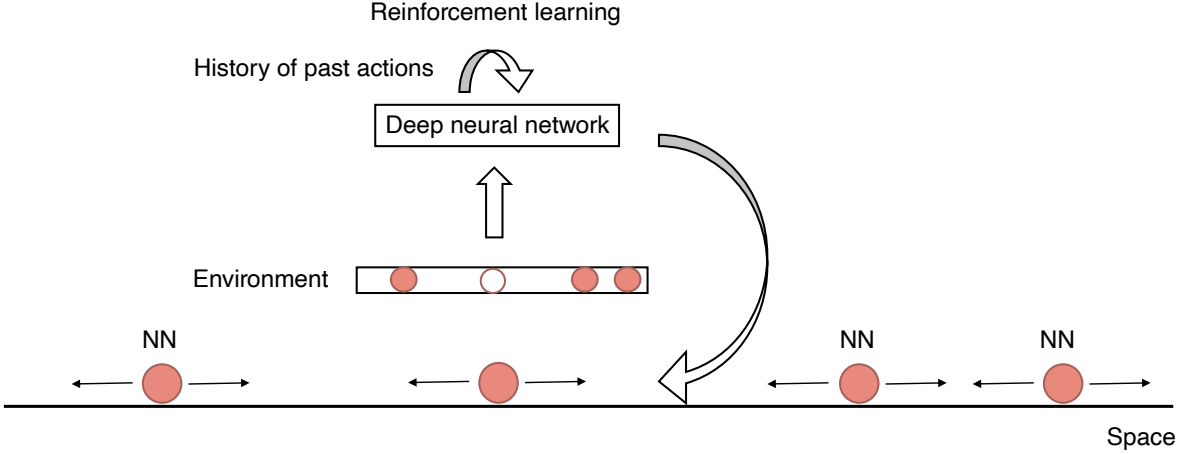


Figure 3.8.: Cartoon of the implementation of a minimal stochastic many-particle system where transition rates are determined by deep neural networks trained using reinforcement learning.

cast in the form of a master equation,

$$\partial_t P(\{x_j\}, t) = \sum_i [f(\{x_i\}, \theta_i) P(\{x_i - 1\}, t) + (1 - f(\{x_i\}, \theta_i)) P(\{x_i + 1\}, t) - P(\{x_i\}, t)] , \quad (3.11)$$

where $P(\{x_j\}, t)$ is the joint probability of finding the particles at lattice position x_j at time t , $P(\{x_i \pm 1\}, t)$ represents the probability of finding particle i at position $x_i \pm 1$, and the rate of moving right $f(\{x_i\}, \theta_i)$, determined by the output of the actor neural network, depends on the state of the system at time t and the set of neural network parameters $\theta_i = (\mathbf{w}_i, \vec{b}_i)$.

Lastly, to determine the temporal evolution of the neural networks, it remains to specify the reward function. In order to emphasize the role of macroscopic feedback in the dynamics of the system, we chose a reward function that is maximized by a collective strategy in the deterministic limit. Specifically, we defined the reward function as

$$r_i(t) = \begin{cases} 1 & x_i \neq x_j, \quad \forall j \neq i, \\ -100 & \text{otherwise,} \end{cases} \quad (3.12)$$

where $r_i(t)$ represents the reward collected by particle i at time t . Our reward function penalizes colliding particles with a large negative reward whilst rewarding the avoidance of such events with a small positive reward. Nevertheless, reinforcement learning problems have been shown to be robust against changes in the magnitude of the rewards [52, 53, 103], so we expect our main results to be also robust against changes in the values assigned to positive and negative rewards.

As the goal of reinforcement learning is to maximize the expected return, defined as the discounted cumulative reward (Equation 3.4), our choice of the reward function corresponds to training the particles to minimize the expected number of collisions. In the deterministic limit of the dynamics, this reward function is maximized by a system in which all particles move in the same direction with constant velocity. In the following sections we set out to characterize the dynamics that emerge from the interplay between the macroscopic dynamics and the optimization of the reward. In particular, we will define spatio-temporal fields to describe the macroscopic dynamics and study their role in the remodelling of the microscopic components of the system, corresponding to the weights and biases of the neural networks.

Before proceeding further, for the sake of clarity we identify the elements of a reinforcement learning problem with the corresponding components of our model

- **Agents:** The actor networks assigned to individual particles.
- **Environment:** The configuration of the lattice at time t , described by the particle positions x_i .
- **Actions:** The action set is given by \mathbb{Z}_2 , with the elements of this field corresponding to the two directions a particle can take.
- **Observations:** The input of the neural networks of individual particles. Corresponds to a neighbourhood centered at the particle position $\mathcal{B}_r(x_i)$.
- **Rewards:** Particles are penalized when colliding and rewarded otherwise.

3.2.1. Details of the numerical implementation

After having defined our theoretical model in the previous section, we next discuss the details of its implementation in the form of numerical simulations. We implemented our model in Python using MxNet [141], an open-source state-of-the-art deep learning library. MxNet offers a flexible high-level computing environment ideally suited to the task at hand.

The stochastic dynamics described by Equation 3.11 are simulated using Gillespie’s algorithm [73] with sequential updating for the length of an episode (corresponding to $t = 20$), at the end of which the neural networks are updated using an inertial second order gradient descent algorithm (ADAM, Equation 3.10) to optimize the loss function. The reaction rates of each particle are determined by an actor-critic scheme composed of two feed-forward deep neural networks with a single hidden layer with $n = 5$ units and tanh activation function.

Explicitly, the reaction rates read

$$\mathcal{R} = (f(\{x_i\}, \theta_1), 1 - f(\{x_i\}, \theta_1), \dots, f(\{x_i\}, \theta_N), 1 - f(\{x_i\}, \theta_N)), \quad (3.13)$$

where $f(\{x_i\}, \theta_i)$ is the probability of particle i moving right given the current state of the system and the neural networks. The size of the input layer is determined by the dimensionality of the partial observation of the state of the system that is provided to every particle. In particular, in our simulations each particle has access to the occupation state of a neighbourhood of size $l = 5$ centered at its position, corresponding to input layers of the same size.

The system is initialized at $t = 0$ drawing the neural network parameters θ from a Gaussian distribution $\mathcal{N}(\mu = 0, \sigma = 0.04)$ and the initial positions of the particles from the set $\{i\}_{i=0}^L$ without replacement, thus ensuring non-overlapping initial positions. Simulations were then run for a total time of $T = 15000$ episodes, at the end of which the state of the system was sampled. Neural networks were trained using ADAM [140] with a learning rate $\eta = 5 \cdot 10^{-4}$. Additionally, networks were sampled more sparingly, their state being saved every 200 episodes.

Actor-Critic Algorithm

We implemented an asynchronous actor-critic algorithm to approximate the policy and the value functions following [104]. Actor-critic algorithms are a classical [51] Reinforcement learning approach where the actor is used to approximate the policy function whereas the critic provides an estimate of the value function (Figure 3.9). In particular, we use neural networks as function approximators for both the policy and the value functions.

This scheme is defined by the loss functions of the actor and the critic networks. The loss of the actor corresponds to the difference between the expected return for a given policy and the optimal value of the return predicted by the critic whereas the loss function of the critic is the difference between the actual return and the predicted return. In compact notation the update rule of the parameters (θ, θ_v) can be written as

$$\nabla_{\theta'} \log \pi(a_t | s_t, \theta') A(s_t, a_t, \theta, \theta_v), \quad (3.14)$$

where $A(s_t, a_t, \theta, \theta_v) = \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma_k V(s_{t+k}, \theta_v) - V(s_t, \theta_v)$ is an estimate of the advantage function [104], quantifying the difference between the estimated state-value and action-state value functions. In the following sections we characterize the dynamics of the simulations at

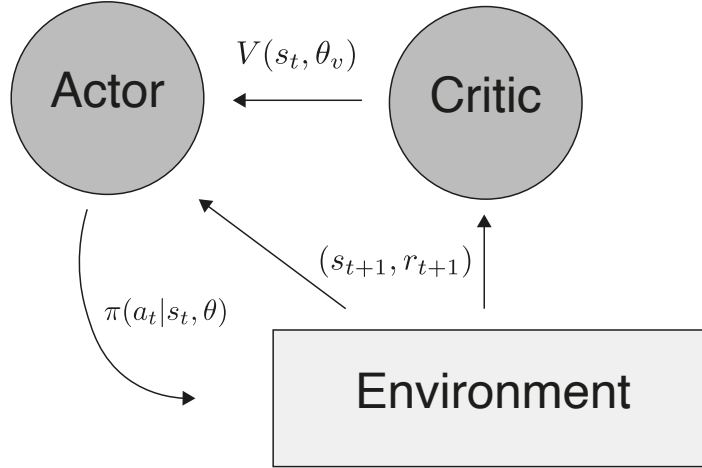


Figure 3.9.: Schematic representation of the basic architecture of an actor-critic algorithm. The actor network approximates the policy function based on the observation from the environment in addition to the approximation to the value function provided by the critic.

the micro and macroscopic levels. Unless otherwise specified, all the results presented in the upcoming sections correspond to systems of $N = 80$ particles in lattices of size $L = 200$.

3.3. Collective dynamics emerge as a result of interactions induced by the shared environment

The stochastic model defined by Equation 3.11 in addition to the neural network dynamics represents a system where particles are trained to maximize individual expected returns. From the perspective of an optimization problem, the individual optimal strategies do not need to result in the emergence of a collective state. Therefore, it is remarkable that the system exhibits collective dynamics as a result of the implicit interactions induced by the shared environment. In this section we qualitatively describe the dynamics of our numerical simulations, simultaneously setting the ground for their rigorous quantification in upcoming sections. To begin, we focus on the behaviour of the system for $\gamma = 0$, where γ is the discount parameter (Equation 3.4).

First, in order to validate our approach, we calculate the average reward obtained by the particles over time (Figure 3.10a). Starting from randomly chosen initial conditions, the average reward evolves non-monotonically before reaching the steady-state value. The reward peaks at $t \approx 2500$ episodes and then decreases monotonically until the steady-state value is reached. However, although the steady-state reward is not optimal, the improvement in the steady-state average reward compared to the initial average reward reveals that particles learn a strategy that performs better than diffusion as a result of macroscopic feedback.

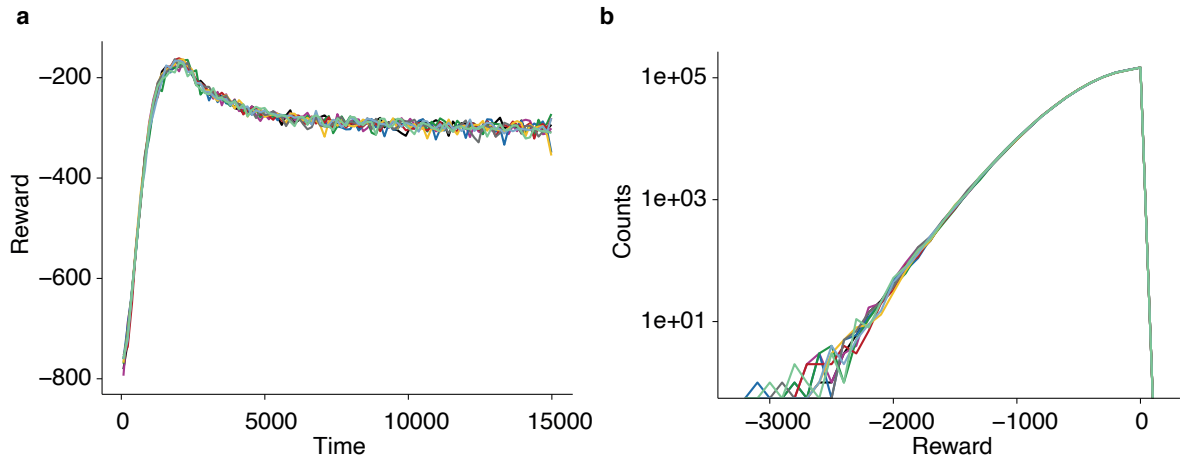


Figure 3.10.: **a)** Average reward over time for nine individual simulations **b)** Distribution of rewards in the steady-state for nine individual simulations. Color represents different simulations.

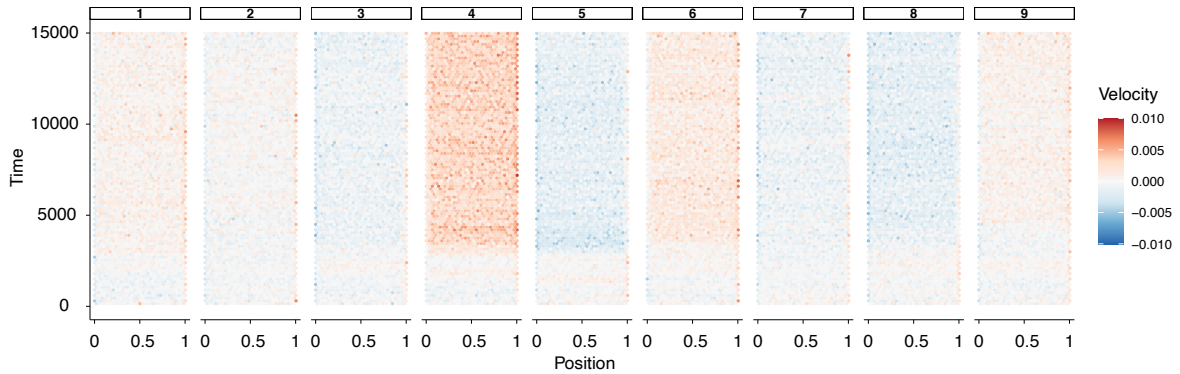


Figure 3.11.: Spatio-temporal evolution of the spatial velocity field $v(x, t)$ for different numerical simulations (columns).

The distribution of rewards in the steady state does not exhibit significant differences across simulations (Figure 3.10b).

We first characterise the population level dynamics by the position and velocity of single particles over time. Additionally, in order to quantify collective effects, we also define the spatial velocity field representing the average velocity at each lattice position over time. We define the trajectory \mathbf{x}_i of particle i as

$$\mathbf{x}_i := \{x_i(t)\}_{t=0}^T, \quad (3.15)$$

where $x_i(t)$ is the position of particle i at time t . The velocity $v_i(t)$ of particle i at time t is defined as

$$v_i(t) := x_i(t) - x_i(t - 1), \quad (3.16)$$

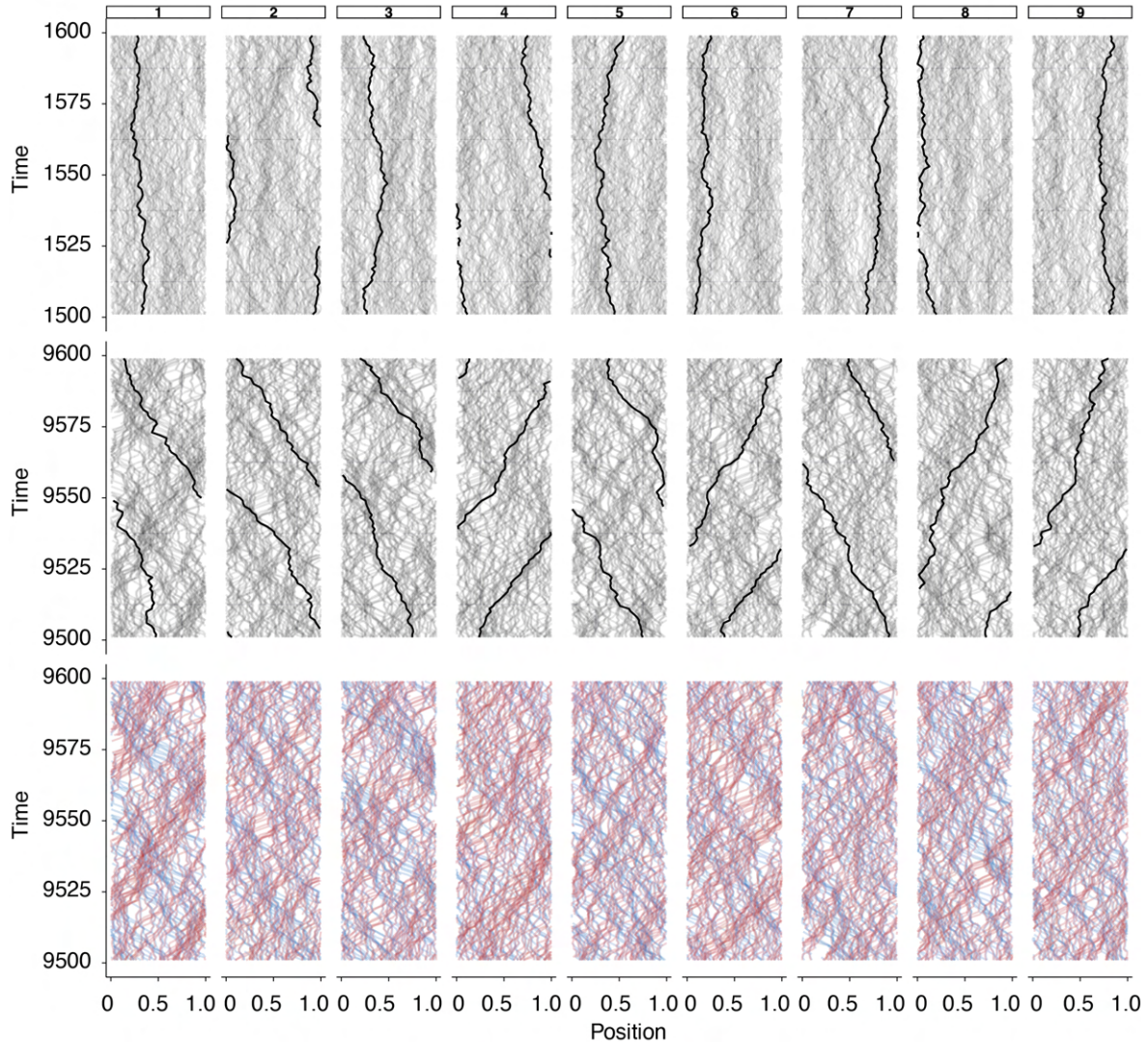


Figure 3.12.: Zoomed-in stochastic trajectories of the system for different numerical simulations (columns) for early (top) and late times (center, bottom). Trajectories of single representative particles are highlighted in black (top, center). Colors represent particles with negative ($v_\infty < 0$) (blue) and positive ($v_\infty > 0$) asymptotic velocity (red).

and finally, we define the spatial velocity field $v(x, t)$ as

$$v(x, t) := \frac{1}{N} \sum_{i=1}^N v_i(t) \delta_{x, x_i(t)}, \quad (3.17)$$

where δ is the Kronecker delta function. In the following we study the evolution of these quantities over time.

We first consider the evolution of the spatial velocity field $v(x, t)$ (Figure 3.11). The results show that the magnitude of the average velocity across the lattice remains small at all times. Therefore, in the following we focus our analysis at the level of single particles in order to discern how the individual dynamics are affected by the population-level feedback.

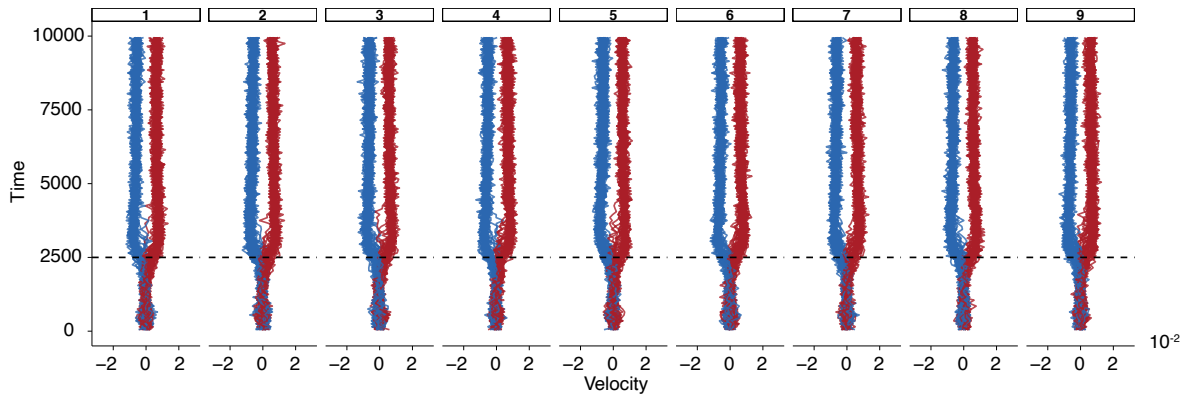


Figure 3.13.: Temporal evolution of velocity distributions for different numerical simulations (columns). Colors represent asymptotic directions of movement. The dashed line indicates the transition time $t_1 \approx 2500$ between unimodal and bimodal velocity distributions.

The trajectories of single particles (Figure 3.12) uncover the existence of two groups of particles moving predominantly in one specific direction revealing that, although not at the population level, the left/right symmetry is broken at the level of single particles.

The symmetry breaking at the particle level is also reflected in the distributions of individual velocities $v_i(t)$ over time (Figure 3.13), showing that after the initial transient regime the asymptotic velocity of single particles, v_∞^i , has a well defined sign, indicating that a given particle asymptotically obtains a finite, constant average velocity.

Additionally, the system exhibits giant number fluctuations (Figure 3.14a), meaning that the fluctuations in the number of particles ΔN in a region of size l grow faster with the average number of particles \bar{N} in such a region than the equilibrium prediction, $\Delta N \propto \sqrt{\bar{N}}$.

To elucidate whether there is a feedback mechanism involved in the formation of the two groups of particles, we calculated the statistical properties of the number of particles belonging to each of the groups and compared our results with the predictions of a random binomial model with parameters $N = 80$ and $p = 1/2$ (Figure 3.14b). Our results reveal that, although the average number of particles per group agrees with the prediction of the null model, the variance of the number of particles belonging to each group is strongly reduced, $\sigma^2 = Np(1-p)$, as corroborated by an F-test for the comparison of the variances. Therefore, this result in combination with the previous observations reveals the existence of a feedback mechanism balancing the ratio of particles in the left and right moving groups.

Nevertheless, considering that individual particles are trained to minimize collisions this strategy is rather counter-intuitive. Two groups of particles moving in opposite directions in a periodic lattice are bound to collide at a rate proportional to their relative velocity $\Delta v = v_+ - v_-$. Therefore, in the following we compare the performance of this strategy to

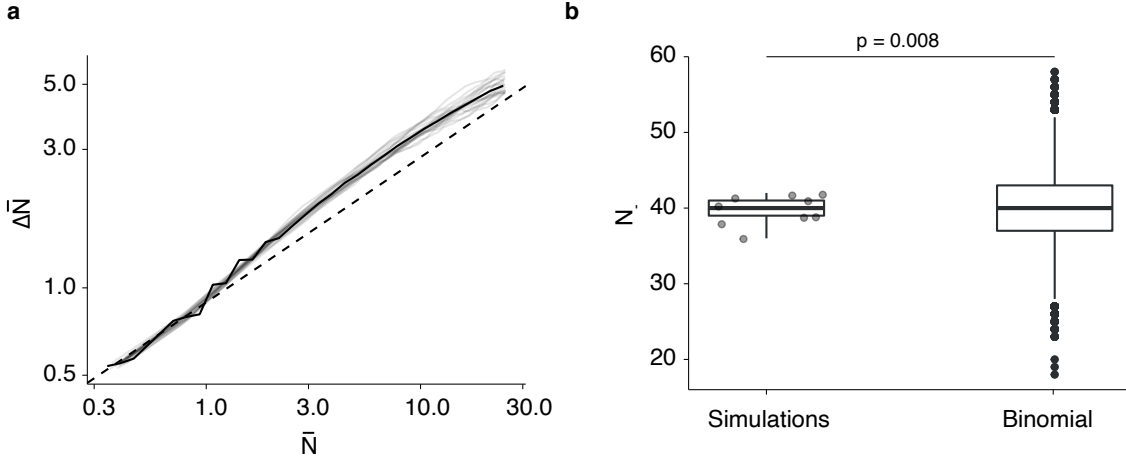


Figure 3.14.: **a)** Giant number fluctuations in the steady state ($t > 14000$). The dashed line corresponds to $\Delta\bar{N} = \sqrt{\bar{N}}$. **b)** Number of particles with positive asymptotic velocity for simulated data and a binomial model with $N = 80$ and $p = 1/2$. P-value corresponds to an F-test performed on the variances of the two distributions.

the optimal strategy in the deterministic limit, corresponding to particles moving in the same direction $\Delta v = 0$.

For the purpose of this comparison, we consider a minimal stochastic model of N particles moving with fixed velocity, i.e. following a deterministic strategy, in a one dimensional lattice of size L and periodic boundary conditions and we study two limiting cases. We consider a system composed by particles moving in the same direction ($\Delta v = 0$) and a system composed of two subpopulations of $N/2$ particles with opposite velocities. The dynamics are simulated using Gillespie's algorithm [73] with constant, identical reaction rates for each particle, $T_{\max} = 1000$ and 10^5 different initial conditions randomly distributed across the lattice.

In order to test whether the stochasticity of the dynamics can destabilise the deterministic optimal strategy, we calculated the empirical distributions of collisions (Figure 3.15) describing the probability of observing a total of N collisions during the simulation. No significant differences were observed in the average collision probability between the two compared scenarios, indicating that the stochasticity of the dynamics can indeed destabilise the deterministic optimal strategy and lead to the emergence of non-trivial collective dynamics.

In summary, our results reveal left/right symmetry breaking at the level of individual particles. Single particles move with a finite, constant asymptotic velocity for $t > 2500$, as reflected by the stochastic trajectories and the dynamics of the velocity distributions. Following the symmetry breaking, the system organises into two equally sized groups of particles. The number of particles in each of two groups is highly regulated, uncovering the

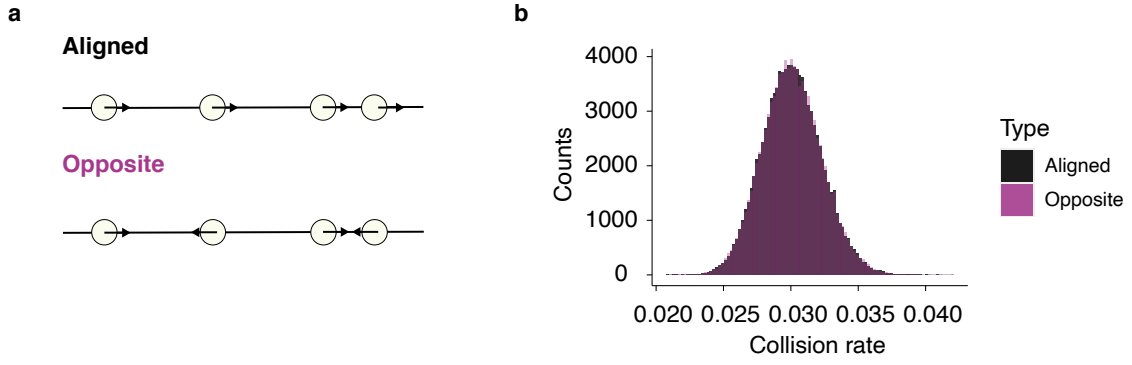


Figure 3.15.: **a)** Cartoon representing the two models under comparison. **b)** Empirical distribution of collision rates.

existence of a feedback mechanism regulating the ratio of particles with positive and negative asymptotic velocity. Furthermore, we compared the counter-intuitive strategy followed by the particles to a system where all particles move in the same direction. The collision probabilities for the two strategies do not show any significant differences, revealing how stochastic effects can lead to destabilisation of the deterministic optimal strategy.

In the following section we characterise the structure of the two groups at the macroscopic level, focusing on both dynamical and steady-state properties.

3.4. Quantification of the dynamics at the macroscopic level

Our numerical results show that trained particles divide into two groups moving ballistically in opposite directions. In the following, we set out to characterise the structure of these groups and their dynamics. With this purpose in mind we define the mean-squared displacement (MSD) as

$$\text{MSD}(\Delta t) = \frac{1}{N} \frac{1}{T} \sum_{\Delta t=0}^T \sum_{i=1}^N (x_i(t_0) - x_i(t_0 + \Delta t))^2, \quad (3.18)$$

where $x_i(t)$ represents the position of particle i at time t .

The mean-squared displacement as a function of the lag time (Figure 3.16) reveals two different dynamical regimes. Before the characteristic time-scale of left/right symmetry breaking ($t_1 \approx 2500$) the dynamics of individual particles are diffusive, as indicated by the linear scaling of the mean-squared displacement $\text{MSD}(\tau) \propto \tau$. On the other hand, particles moves ballistically for $t > t_1$, reflected in the scaling $\text{MSD}(\tau) \propto \tau^2$. Having characterised the dynamics of individual particles, we then set out to investigate the left/right symmetry breaking by quantifying the evolution of spatial observables of the dynamics.

First, we focus on the statistics of the distribution of particles across the lattice over time.

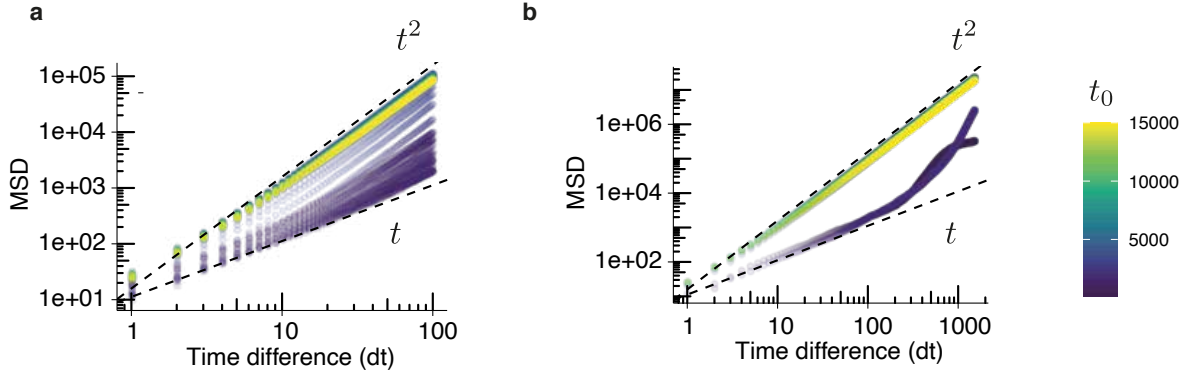


Figure 3.16.: **a)** MSD as a function of the time difference dt for time windows of size 150 episodes. **b)** MSD as a function of the time difference dt for time windows of size 1500 episodes. Color represents the starting episode of the time window.

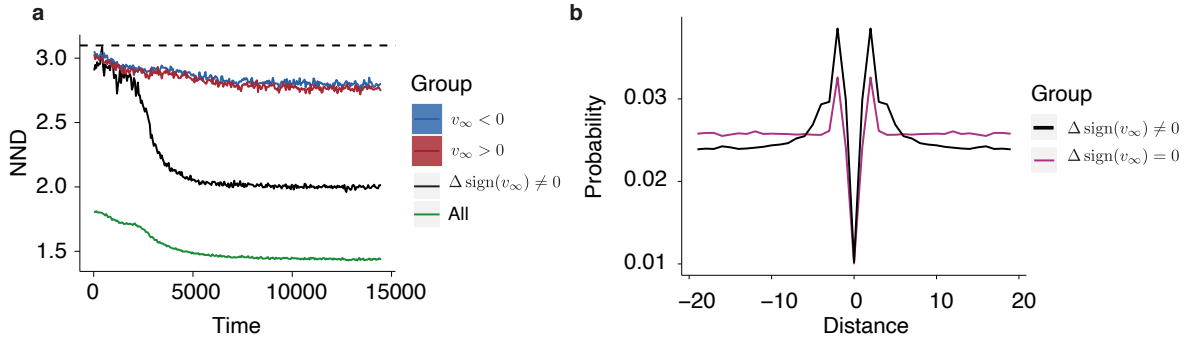


Figure 3.17.: **a)** Nearest neighbour distance for particles with $v_\infty < 0$ (blue), $v_\infty > 0$ (red), particles moving in opposite directions ($\Delta \text{sign}(v_\infty) \neq 0$, black) and all particles (green). The dashed line corresponds to the predicted nearest neighbour distance for $N = 40$ particles randomly distributed in a lattice of size $L = 200$. **b)** Occupancy probabilities given by Equation 3.21 for particles moving in the same direction ($\Delta \text{sign}(v_\infty) = 0$, purple) and moving in opposite directions ($\Delta \text{sign}(v_\infty) \neq 0$, black) as a function of the relative distance.

We tested whether the spatial distribution of particles depends on the relative velocity of the groups under consideration. In particular, we calculated the non-overlapping average nearest neighbour distance (NND) as a function of time. This quantity measures the average distance to the nearest occupied lattice site and is defined as

$$\text{NND}(t) = \frac{1}{N} \sum_i \min_{j, x_i(t) \neq x_j(t)} |x_i(t) - x_j(t)|, \quad (3.19)$$

where the sum is taken over all particles in the system and the minimum over particles moving in the same or opposite directions.

The NND dynamics (Figure 3.17a) indicate that the average distance between particles decreases over time, remaining smaller than expected by chance. Furthermore, Figure 3.17a also shows that particles moving in opposite directions are on average closer than particles

moving in the same direction.

In order to compare the steady-state statistical properties of particles moving in the same and opposite directions, we compute the probability of lattice sites being occupied as a function of the distance from a reference particle from the simulation data. To calculate this quantity, we define the normalized centered density profiles as

$$\rho(x, t) = \frac{1}{N^2} \sum_{i,j} \delta_{(x_j(t)-x_i(t),x)}, \quad (3.20)$$

where the sum is performed over all particles, $x_i(t)$ is the position of particle i at time t and δ represents the Kronecker delta function. The normalized centered density profile $\rho(x)$ is non-zero only for those values of x that correspond to the distance between a pair of particles. In turn, the probability of neighbouring sites being occupied in the steady-state is given by the time average of the normalized centered density profile

$$\bar{\rho}(x) = \frac{1}{T} \sum_{t=t_0}^T \rho(x, t). \quad (3.21)$$

Additionally, we remove the contribution from the reference particles from the density profiles in order to retain only the contributions to the occupancy probability coming from distinct particles. The occupancy probabilities (Figure 3.17b) reveal that particles are in closer proximity of particles moving in the opposite direction than of particles moving in the same direction. Furthermore, the drop in the occupancy probability at $\Delta x = 0$ indicates that particles preferentially do not overlap, irrespective of their direction of movement.

Put together, our results show a decrease in the inter-particle distance over time, with neighbouring sites being preferentially occupied by particles moving in the opposite direction. These findings reveal the emergence of spatial clusters, indicating that the density of particles across the lattice does not remain constant. Additionally, the system exhibits anti-ferromagnetic order, as shown by the occupancy probabilities (Figure 3.17b).

In the next section we will study how these findings are reflected at the microscopic level. In particular we will quantify the temporal evolution of the neural network parameters and the response of the networks to different inputs.

3.5. Characterization of the neural network dynamics

The microscopic scale of the system corresponds to the parameters of the neural networks, representing the multiple degrees of freedom of each particle. These parameters, together

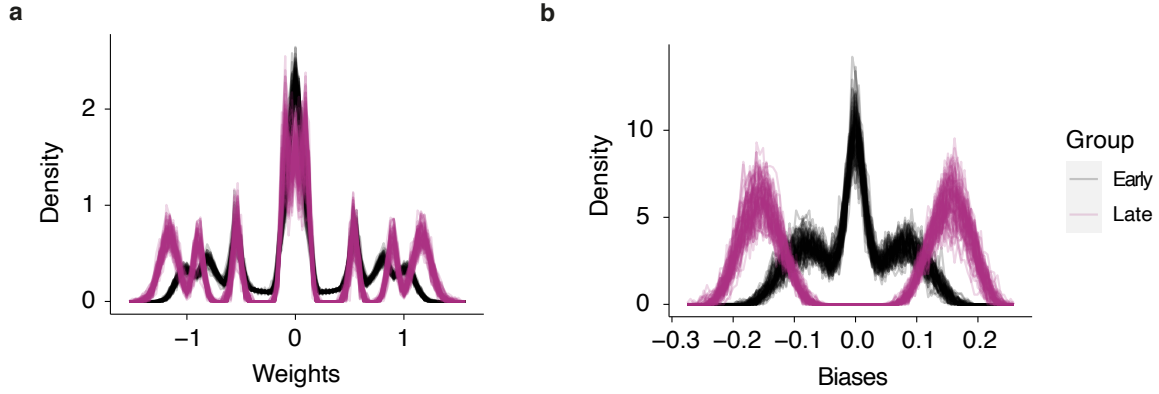


Figure 3.18.: Actor networks early and late-time distributions of **a**) weights and **b**) biases.

with the activation function, define the mapping between the microscopic and macroscopic scales (Equation 3.5). In this section we quantify the changes that these parameters undergo over time and how that in turn is reflected in the output of the neural networks.

First, in order to evaluate the impact of learning on the structure of the actor networks we compared the distribution of neural network parameters before and after t_1 (Figure 3.13), the transition time between unimodal and bimodal velocity distributions. The dynamics were divided in two regimes that we termed early times ($t < 2500$) and late times ($t > 7500$), corresponding to the learning phase and the asymptotic limit of the dynamics respectively.

Both weights (Figure 3.18a) and biases (Figure 3.18b) exhibit similar changes in their distribution across particles as a result of training, as reflected by the transition from a unimodal distribution at early times to a multi-modal distribution at late times. However, whereas the late-time weights distribution displays a mode centered at 0, the biases distribution does not, the two modes of this distribution being located at non-zero values. The results indicate that neural networks become more specialized after training, as reflected by the increase in the heterogeneity of the parameter distribution.

The relationship between input and output and its dependence on the network parameters is obscured by the complex architecture of the neural networks. However, it is precisely the output of the neural networks what determines the temporal evolution of the system and in turn leads to the remodelling of the neural networks. Accordingly, in order to understand the functional effect of the changes in parameter distributions over time, we calculated the temporal evolution of the output of the actor neural networks for several input signals, corresponding to the probability of moving in one direction given an observation.

In order to systematically study the impact of different observation signals in the output of single networks we calculated the two-dimensional UMAP representation of the probability $P(x \rightarrow x + 1 \mid s_t)$ of having an instantaneous positive velocity for a total of 126 input signals,

corresponding to all possible observations with by five or less particles. UMAP [142] is a non-linear dimensionality reduction method that preserves neighbourhood relationships during the low dimensional embedding. The algorithm takes the probability of having a positive velocity given the environmental state for individual neural networks at specific time points as input and returns the coordinates of the embedding in the two dimensional UMAP space. Although the UMAP coordinates cannot be easily interpreted in terms of the original variables, the algorithm conserves the topology of the original high-dimensional data and the results can be interpreted by over-laying additional information to the UMAP representation. Therefore, UMAP provides a suitable method for reducing the dimensionality of the observation space.

The two-dimensional UMAP representation (Figure 3.19) shows the emergence of two groups of particles, corresponding to particles with positive and negative asymptotic velocity, that do not exhibit significant differences in the instantaneous average reward across the system. The results reveal a initial phase characterised by movements between discrete clusters in UMAP space (Figure 3.19a), followed by the evolution in a single connected cluster after t_1 . Additionally, the branches corresponding to particles with positive and negative velocity (Figure 3.19) contain particles with both positive and negative biases (Figure 3.19d), indicating that there is no connection between the asymptotic velocity of the network and the average bias of the corresponding neural network. Trajectories of single particles provide insight on the intra-cluster dynamics (Figure 3.19e) and do not exhibit significant differences between simulations (Figure 3.19f).

We hypothesized that the observed symmetry-breaking could be the result of diverging evaluations of the same observations during the learning phase. In order to test this hypothesis we evaluated the expected instantaneous reward for all environments occupied with five or less particles based on the actor networks output and projected the data using UMAP to study its low dimensional structure.

As previously, the graphical representation of the UMAP embedding (Figure 3.20) shows how particles progressively separate into two groups, reflecting the left/right symmetry breaking at the microscopic level. The dynamics are again characterised by an initial phase where particles move between discrete clusters in UMAP space, followed by a regime where all particles move in a single connected cluster (Figure 3.20a). The lack of differences in the estimated reward between the two groups (Figure 3.20c) indicates that the symmetry breaking does not result from differential evaluation of the same observations during training. Additionally, the composition of the branches corresponding to particles with positive and negative asymptotic velocity (Figure 3.20d) does not depend on the average bias of the corresponding neural

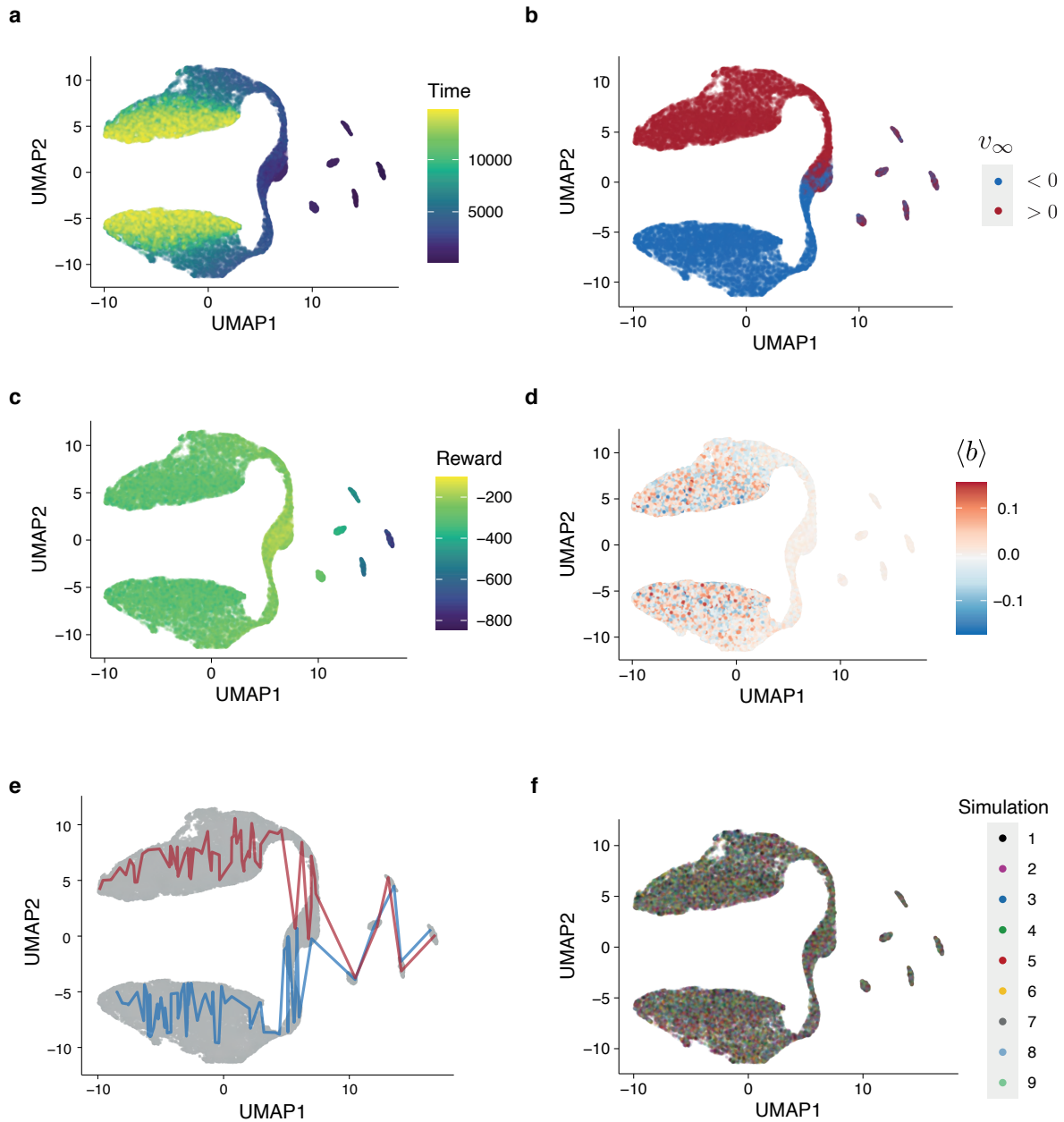


Figure 3.19.: UMAP representation of the output of the actor neural networks for all observations with by five or less particles. Color indicates **a)** time (in units of episodes), **b)** asymptotic velocity of individual particles, **c)** instantaneous average reward across the system, **d)** mean bias of the individual neural networks, **e)** trajectories of selected particles with $v_\infty < 0$ (blue) and $v_\infty > 0$ (red), **f)** different realizations of the simulations.

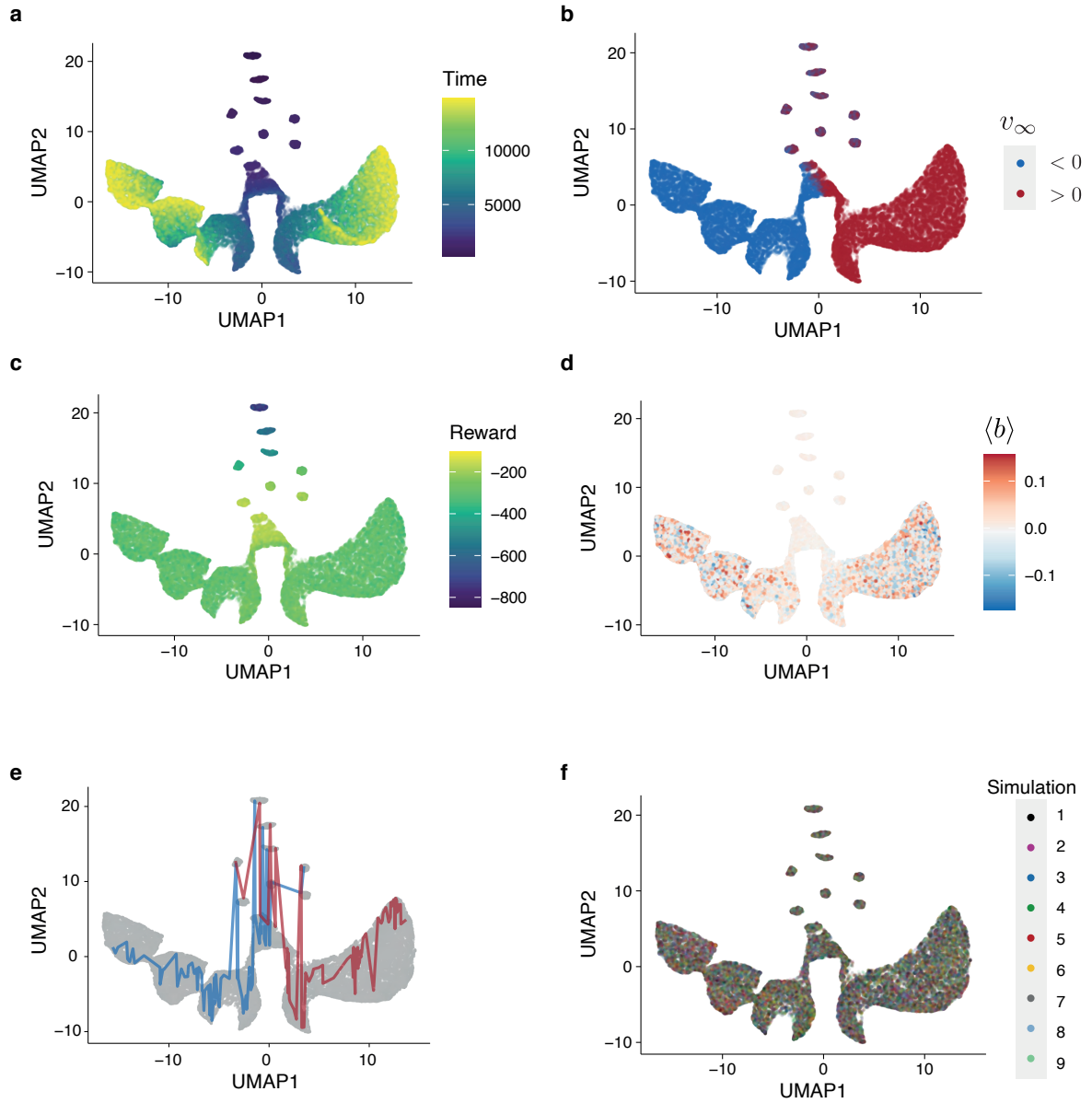


Figure 3.20.: UMAP representation of the expected instantaneous reward for all observations with by five or less particles. Color indicates **a)** time (in units of episodes), **b)** asymptotic velocity of individual particles, **c)** instantaneous average reward across the system, **d)** mean bias of the individual neural networks, **e)** trajectories of selected particles with $v_\infty < 0$ (blue) and $v_\infty > 0$ (red), **f)** different realizations of the simulations.

network and the trajectories of individual particles (Figure 3.20e) do not show significant differences between simulations (Figure 3.20f).

Together, our results show how neural networks are dynamically remodelled over time. First, we have shown that the distributions of weights and biases (Figure 3.18) become more heterogeneous after the time-scale associated with the left-right symmetry breaking. Additionally, we have explored the input-output connection using a low-dimensional representation of the output of the neural networks for a large set of observation signals. Our results (Figure 3.19) reveal the existence of two clusters in UMAP space, corresponding to particles with positive and negative asymptotic velocity. The two clusters are composed of particles with positive and negative biases, indicating that there is no connection between the average bias of a neural network and the asymptotic velocity of the corresponding particle. Finally, we quantified whether the left/right symmetry breaking was caused by differential evaluation of the same environments during training. The results (Figure 3.20) reveal that there are no significant differences in the evaluation of the expected reward during training for particles with positive and negative asymptotic velocities.

In the next section we study the interplay between the microscopic and macroscopic levels and show that it leads to the emergence of complex inter-particle interactions.

3.6. Emergence of complex interactions as a result of the interplay between the macroscopic and the microscopic scales

One of the main features of deep reinforcement learning tasks compared to supervised or unsupervised learning is the lack of a training data set, as the training data is dynamically generated by the experiences of the agents during training. Therefore, correlations in the dynamics of the system can have strong effects in the reward collection and, in turn, in the remodelling of the neural networks. In the following, we explore whether and how the interplay between macroscopic and microscopic processes gives rise to interactions between particles.

In particular, we explicitly calculated the probability of moving with positive instantaneous velocity $P(x \rightarrow x + 1 \mid s_t)$ for individual particles over time. The results (Figure 3.21) reveal that particles dynamically adapt their response to different observations, in a process that takes place over several time scales, as reflected by the evolution of $P(x \rightarrow x + 1 \mid s_t)$. Whereas initially the probability of moving left or right is the same for all input signals, particles learn to classify different observations, quickly reaching a stationary value for certain input signals

whereas in other cases $P(x \rightarrow x + 1 \mid s_t)$ is continuously updated during the dynamics.

Figure 3.21 reveals the emergence of complex inter-particle interactions during the dynamics. First, particles quickly learn to classify signals with left and right nearest neighbours based on the difference between the number of left and right nearest neighbours (Figure 3.21a-c). Therefore, the system develops short-range repulsive interactions for time-scales smaller than the left-right symmetry breaking time-scale t_1 . Environments without nearest-neighbours are classified more slowly (Figure 3.21b-d), as in this case no collision takes place after a single stochastic step irrespective of the direction of movement. These environments are later on classified depending on higher order features (Figure 3.21a-c), leading to the emergence of attractive long-range interactions.

The analysis of the evolution of the output of individual neural networks reveals the emergence of effective interactions. The character of these interactions is modified during the dynamics, reflecting the existence of a hierarchy of timescales associated with the learning of new features of the observation signal. Whereas features like the difference between the number of left and right nearest neighbours are quickly learned, other features are incorporated more slowly into the interactions. Therefore, as a result of the interplay between the microscopic and the macroscopic dynamics, the system exhibits short-range repulsive, long-range attractive interactions in the asymptotic regime.

3.7. Increasing the discount parameter induces a transition between asymptotic dynamical regimes

In previous sections we have characterized the dynamics of the system for $\gamma = 0$. The discount γ (Equation 3.4) is one of the central parameters of any RL task, as it introduces correlations in the calculation of the loss function. In fact, in practical applications with long-term goals large discount factors ($\gamma > 0.9$) are commonly chosen [53, 105]. Although a finite discount parameter can give rise to more complex dynamics, it also renders analytic descriptions of the dynamics unfeasible. In this section, leveraging of our understanding of the $\gamma = 0$ case, we explore the effects of a finite discount parameter in the dynamics of the system, focusing our attention on the macroscopic scale.

First, to gain some intuition we qualitatively describe the macroscopic dynamics as in Section 3.3. In particular, we calculated the spatio-temporal dynamics of the coarse-grained velocity field $v(x, t)$ (Figure 3.22), of the stochastic trajectories of individual particles \mathbf{x}_i (Figure 3.23) and the distribution of velocities $v_i(t)$ across the population (Figure 3.24).

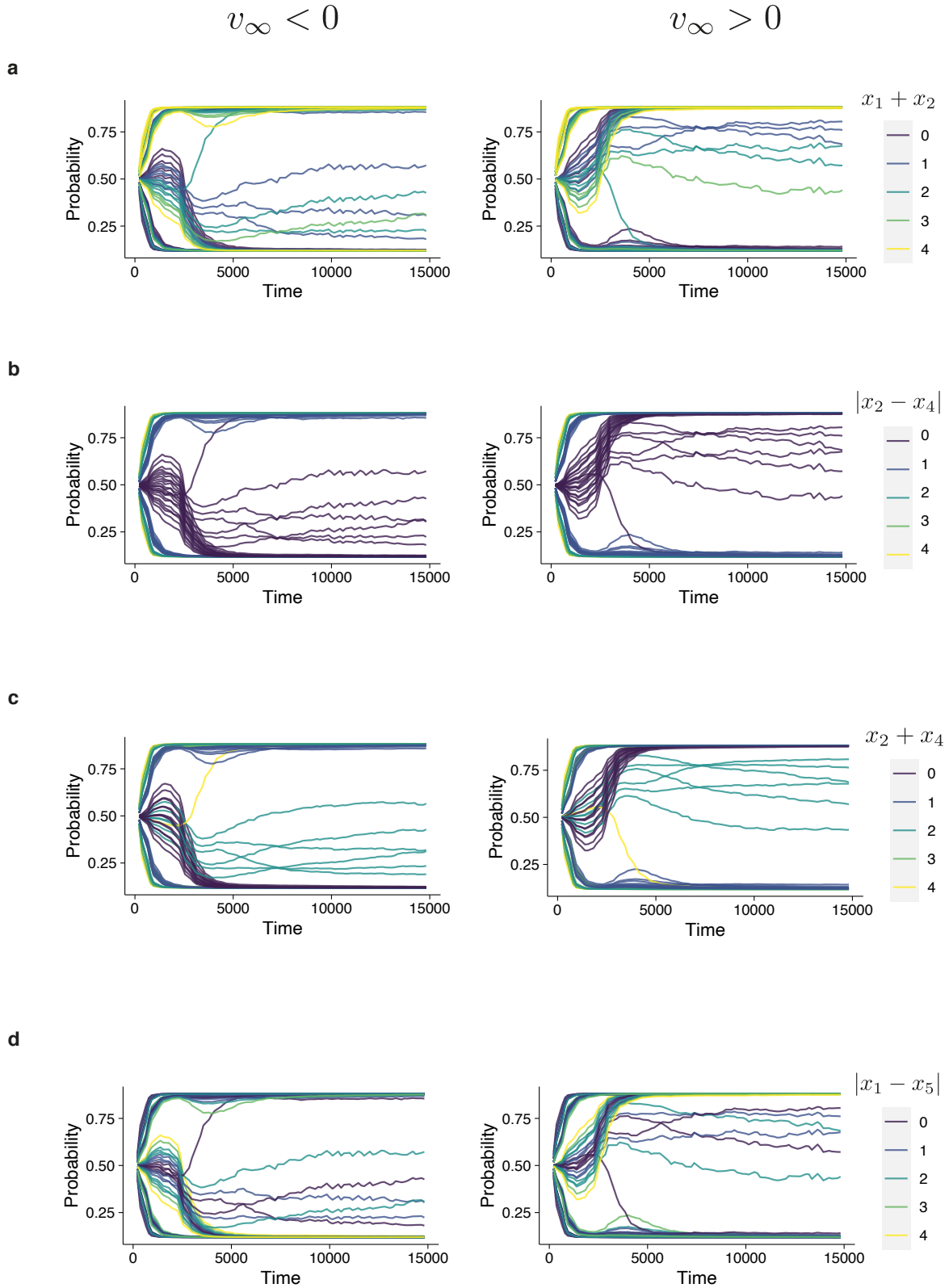


Figure 3.21.: Temporal evolution of the average $P(x \rightarrow x + 1 | s_t)$ for all input signals $s_t = (x_1, x_2, x_3, x_4, x_5)$ with a total of five particles for particles with $v_\infty < 0$ (left) and $v_\infty > 0$ (right). Colors correspond to **a**) number of left neighbours **b**) absolute value of the difference in the number of left and right nearest neighbours **c**) absolute value of the difference in the number of left and right next nearest neighbours.

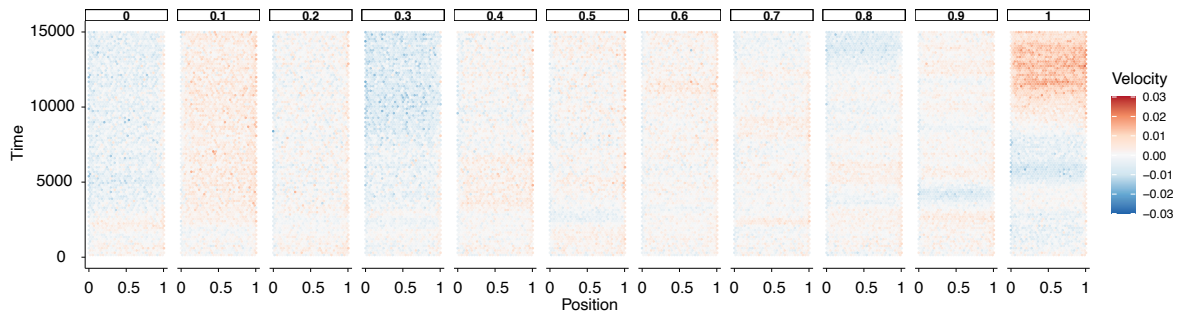


Figure 3.22.: Spatio-temporal evolution of the coarse-grained velocity field for different values of the discount parameter γ .

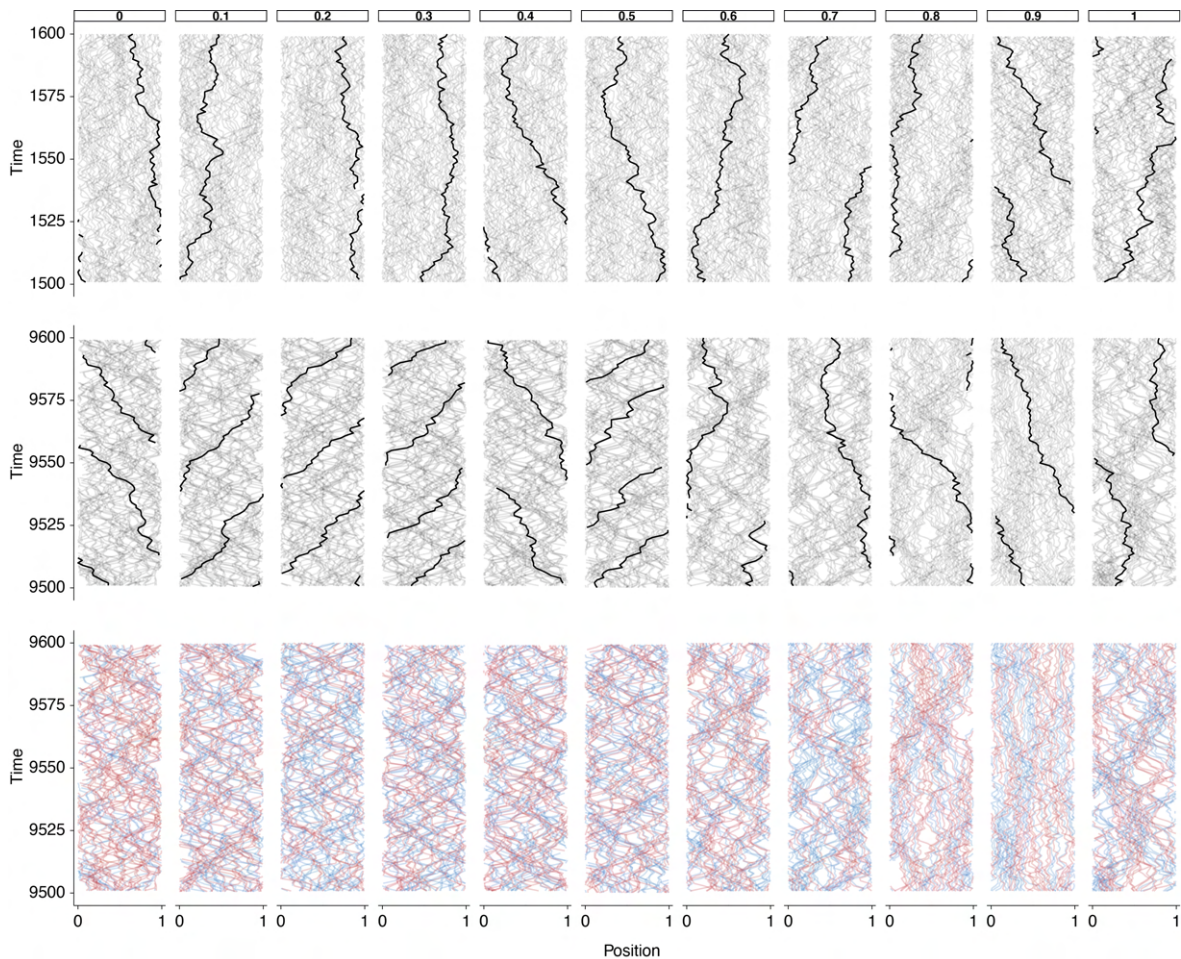


Figure 3.23.: Zoomed-in stochastic trajectories of the system for different values of the discount parameter γ (columns) for early (top) and late times (center, bottom). Trajectories of single representative particles are highlighted in black (top, center). Colors represent particles with $v_\infty < 0$ (blue) and $v_\infty > 0$ (red).

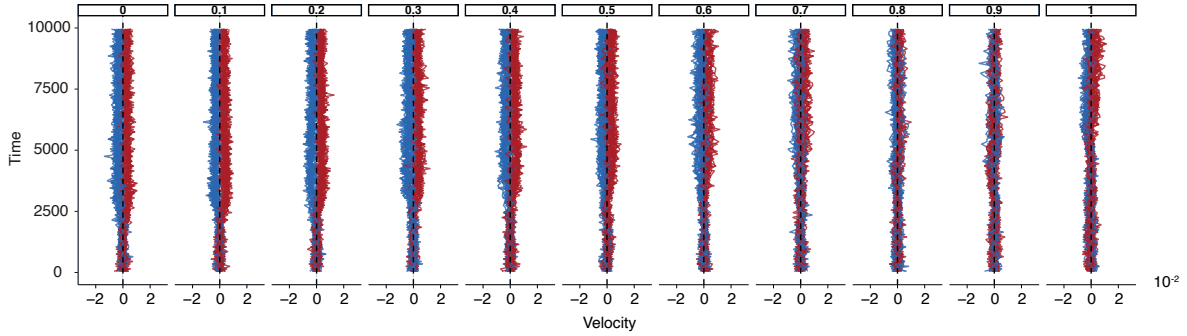


Figure 3.24.: Temporal evolution of velocity distributions for different values of the discount parameter γ . Colors represent particles with $v_\infty < 0$ (blue) and $v_\infty > 0$ (red) .

Although for small values of the discount parameter the results are qualitatively similar to the $\gamma = 0$ case (Figure 3.22, Figure 3.23), exhibiting a cross-over between diffusive and ballistic regimes and left/right symmetry breaking at the level of individual particles, for larger values of the discount parameter the velocity of individual particles undergoes frequent sign changes, as reflected by the stochastic trajectories (Figure 3.23) as well as the velocity distributions (Figure 3.24).

These results point at the existence of two different asymptotic regimes as a function of the discount parameter. On the one hand, for small values of the discount parameter two stable equally-sized groups of particles with opposite signs of the asymptotic velocity form whereas, on the other hand, for larger discount values there is no persistent left/right symmetry breaking at the level of individual particles. In the next section we make use of tools from the field of statistical physics, in particular autocorrelation functions, to explore this change in the dynamics.

3.7.1. Autocorrelation functions of the individual velocities

A fundamental quantity in statistical physics [7], correlation functions describe the co-dependence of two or more fields, providing information about the typical scales over which the fields of interest are correlated. We define the two-point correlation function of fields X and Y as

$$C_{XY}(s, t) = \langle X(s)Y(t) \rangle - \langle X(s) \rangle \langle Y(t) \rangle. \quad (3.22)$$

In particular, if the two fields X and Y are taken to be the same, we refer to the resulting correlation function C_{XX} as the autocorrelation function of the field X . Autocorrelation functions describe the relationship between the values of a field at two different positions.

In order to study the transition between different asymptotic regimes as a function of the discount parameter, we calculated the temporal autocorrelation functions of the velocities of

individual particles, defined as

$$C_{vv}(\Delta t) = \frac{1}{N} \sum_{i=1}^N \langle v_i(t)v_i(t + \Delta t) \rangle - \langle v_i(t) \rangle \langle v_i(t + \Delta t) \rangle, \quad (3.23)$$

where the brackets $\langle \cdot \rangle$ denote time averages and we have assumed the stationarity of the velocity distributions.

To quantify the dependence of the autocorrelations on the discount factor we ran simulations increasing γ from 0 to 1 with a step size of 0.01. The resulting autocorrelation functions are well described by an exponential decay as a function of the time lag Δt (Figure 3.25a). Fitting the correlation functions to an exponential function returns the autocorrelation time, an estimate of the typical timescale over which individual velocities are correlated in the system. In particular, we fit the correlation functions to the following functional form

$$C(\Delta t) \approx e^{-\Delta t/\tau}, \quad (3.24)$$

where τ is the autocorrelation time.

The results of our analysis are presented in Figure 3.25b, revealing a non-monotonic dependence of the autocorrelation time on the discount factor. In particular, the autocorrelation time as a function of the discount exhibits two local maxima located at $\gamma \approx 0.45$ and $\gamma = 1$. The second of these values corresponds to the regime where all future rewards are weighted equally in the calculation of the expected return. In this limit the autocorrelation time of the individual velocities is of the same order of magnitude as the total simulation time, indicating that individual particles always move in the same direction. As all future rewards are weighted equally, changes in the policy followed by individual particles have a large impact in the cumulative reward, meaning that as a result the strategy followed by the particles is persistent in time. On the other hand, for finite discount values the impact of changes in the strategy of individual particles in the cumulative discounted reward decays over time according to the discount factor. The first maxima of the autocorrelation time corresponds to the transition between the two aforementioned asymptotic dynamical regimes. The non-monotonic behaviour of the autocorrelation time around this point indicates a connection between the relevant time-scales of the system: on the one hand the reward time-scale defined by the discount factor and on the other the time-scale of experiences relevant for the neural network remodelling, defined by the episode length.

In summary, in this section we have shown the intricate connection between the discount parameter and the dynamics of the system, as revealed by the autocorrelation time of in-

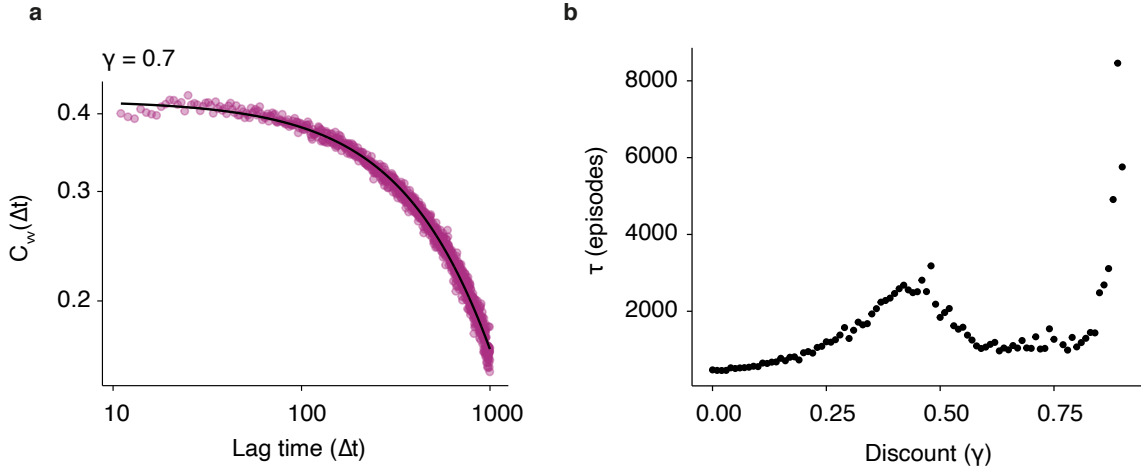


Figure 3.25.: **a)** Example fit of the velocity autocorrelation function to a decaying exponential function $e^{-\Delta t/\tau}$ for $\gamma = 0.7$. **b)** Autocorrelation time τ as a function of the discount parameter.

dividual velocities, with finite discount factors effectively inducing extended interactions in time.

3.8. Effect of demographic disorder on the collective dynamics

Hitherto we have focused our attention on a class of homogeneous systems where all agents share a common reward function. Nevertheless, the presence of individuals with competing goals can have a dramatic effect on the system dynamics and, in particular, on the optimal strategy. In this section we explore the effects induced by the addition of demographic disorder on the system dynamics.

Specifically, we include demographic disorder in our description of the system as a second species of particles with the goal of minimizing, instead of maximizing, the cumulative discounted reward (Equation 3.4). The instantaneous reward function is otherwise the same for both species of particles, given by Equation 3.12. Therefore, the system is effectively formed by two species of particles, one with the goal of minimizing the number of collisions in which they are involved and one with the opposite goal. Additionally, for the purposes of this section we considered particles associated to deep neural networks with three hidden layers composed of five units each. This architecture provides more robust results while qualitatively not affecting the dynamics.

In the following we focus on the stochastic trajectories \mathbf{x}_i of single particles (Figure 3.26)

and the spatio-temporal dynamics of both the density field $\rho(x, t)$ (Figure 3.27), defined as

$$\rho(x, t) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), x)}, \quad (3.25)$$

where $\delta_{(x,y)}$ is the Kronecker delta, and the coarse-grained velocity field $v(x, t)$ (Figure 3.28) to explore the behaviour of the system for different values of the discount parameter γ .

For large values of the discount parameter, the steady state corresponds to a phase of homogeneous density where individual particles move ballistically. The dynamics are characterised by the emergence of two groups of particles with well-defined asymptotic velocities, indicating that the left/right symmetry is broken at the level of individual particles.

However, the dynamics of the system for small values of the discount parameter exhibit strikingly different phenomenology. Specifically, we observe the formation of droplet-like structures (Figure 3.26) composed of a dense core of return-minimizing particles surrounded by a dilute layer of return-maximizing particles.

Motivated by these observations we hypothesize that the formation of aggregates for small values of the discount parameter is related to the temporal correlations induced by a finite discount factor. For small values of the discount parameter particles only the immediate surroundings of a particle are taken into consideration in the calculation of the discounted cumulative reward. Therefore, being in close proximity of collision-seeking particles could prove to be an effective strategy for collision-avoiding particles to maximize the information provided by the environment and minimize the likelihood of colliding.

3.9. Conclusions

In this chapter we have explored how the coupling between microscopic and macroscopic degrees of freedom induced by a shared environment can give rise to collective behaviour in a stochastic many-particle system where the dynamics are defined by deep neural networks trained using reinforcement learning (Figure 3.3). In particular, we have focused on a one dimensional stochastic lattice gas as our minimal model.

We have shown that interacting lattice systems driven by reinforcement learning exhibit non trivial collective behaviour, as reflected by the left/right symmetry breaking at the level of individual particles in the steady state (Figure 3.12). Indeed, we observe that in the steady state particles form two equally sized groups corresponding to the two possible signs of their asymptotic velocities. The formation of these groups relies on a feedback mechanism, as indicated by the observed reduction in the variance of the number of particles in the groups

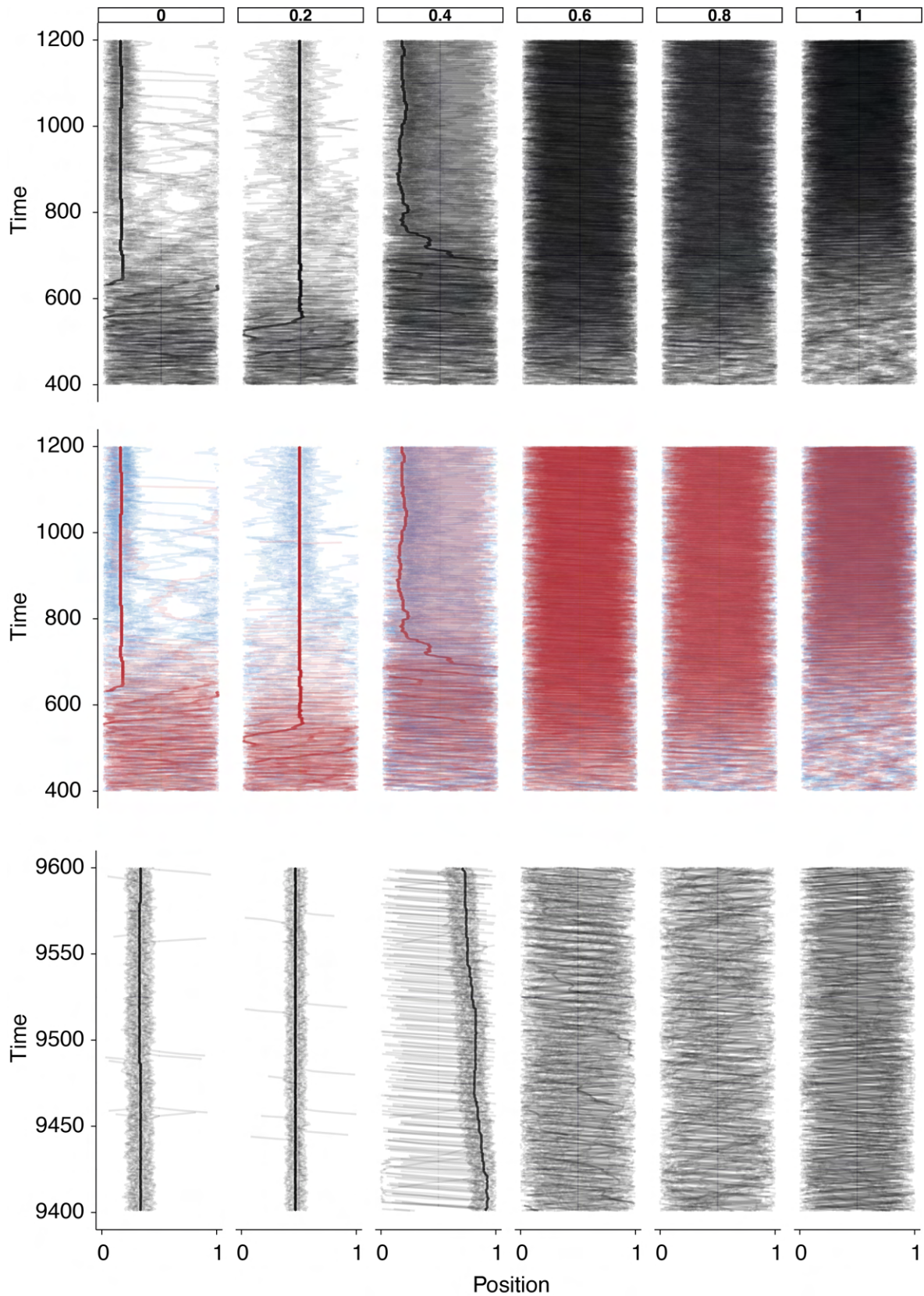


Figure 3.26.: Zoomed-in stochastic trajectories of the disordered system for early (top, center) and late regimes (bottom) and different values of the discount parameter γ . Color (center) corresponds to collision-seeking (red) and collision-avoiding (blue) particles.

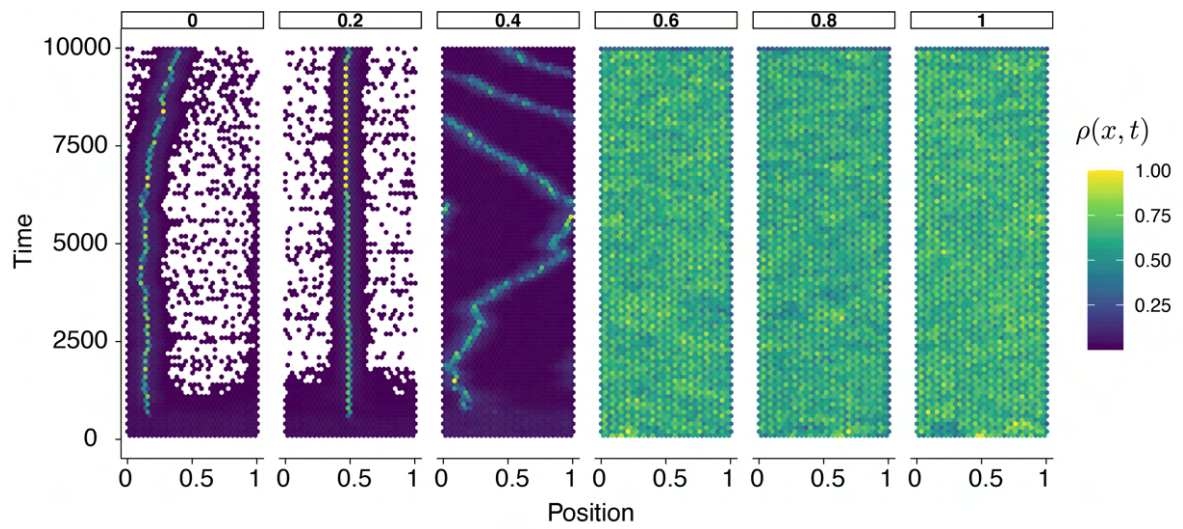


Figure 3.27.: Spatio-temporal evolution of the normalized density field $\rho(x, t)$ for different values of the discount parameter γ for the disordered system.

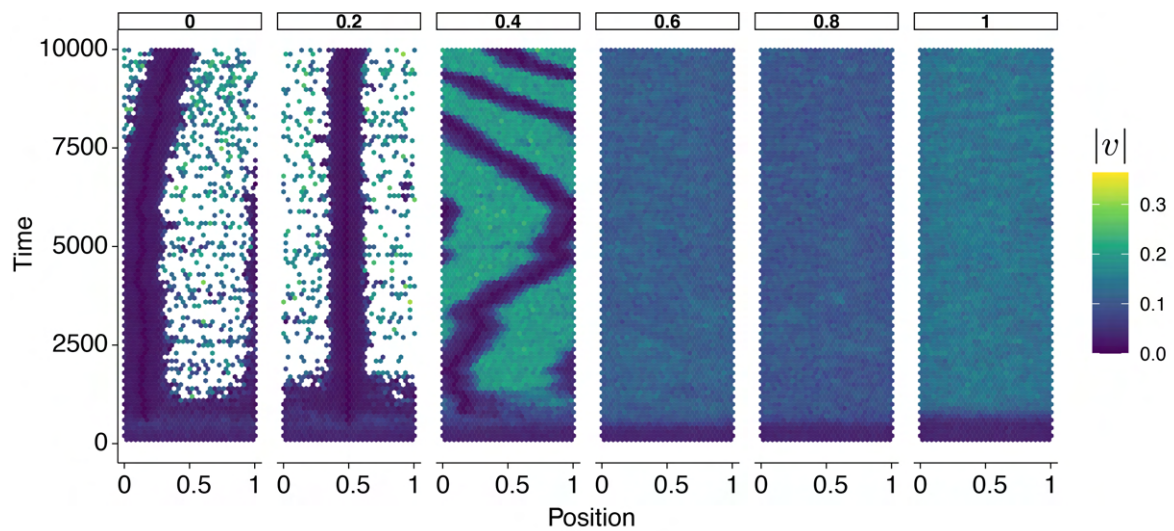


Figure 3.28.: Spatio-temporal evolution of the absolute value of the coarse-grained velocity field for different values of the discount parameter γ for the disordered system.

compared to a null model (Figure 3.14). This counter-intuitive strategy is the result of the stochastic dynamics destabilizing the deterministic optimal strategy $\Delta v = 0$.

At the macroscopic level we observed that the distance between particles decreases over time without overlapping, as reflected by the nearest neighbour distance as well as by the normalized density histograms in the steady state (Figure 3.17).

Moreover, the temporal evolution of the neural network parameters reveals that the networks undergo severe remodelling during the learning phase, as reflected by the trajectories of individual particles in UMAP space, where particles cluster in two groups based on the sign of their asymptotic velocity (Figure 3.19, Figure 3.20).

To complete our discussion, we studied the correlation between the macroscopic and the microscopic degrees of freedom, showing how the emergence of short-range attractive, long-range repulsive interactions is a result of the interplay between these two scales, as illustrated by the temporal evolution of the output probabilities of the neural networks for different input signals (Figure 3.21).

Whereas the results discussed above refer to a homogeneous system without memory, we also studied the impact of memory effects and demographic disorder in the dynamics of the system. In particular, memory effects, controlled by the discount parameter γ , induce a transition between two different asymptotic dynamical regimes as this parameter is modified, resulting on a steady state that depends on the weight assigned to future compared to immediate rewards (Figure 3.23). Specifically, for small values of the discount parameter the system exhibits an asymptotic dynamical regime characterized by left/right symmetry breaking at the level of individual particles, whereas for larger values of the discount, the asymptotic dynamics exhibit frequent changes in the predominant direction of movement of the particles.

Furthermore, the existence of a transition between different asymptotic dynamical regimes is supported by the temporal autocorrelation functions of the individual velocities that decay exponentially over time for all values of the discount parameter, with corresponding autocorrelation times that depend non-monotonically on the discount parameter, peaking at an intermediate value of $\gamma \approx 0.45$ (Figure 3.25).

Lastly, we have shown that the introduction of demographic disorder in the system can lead to the formation of complex spatial structures that are stable over long periods of time. In particular, we introduced demographic disorder in the form of a second species of particles with the goal of minimizing, instead of maximizing, the expected cumulative discounted reward. As a result, we observed for small values of the discount parameter the emergence

of tubular structures formed by a dense core of minimizing particles surrounded by a dilute layer of maximizing particles (Figure 3.26). We hypothesized that this behaviour could result from short-range information propagation for small values of the discount parameter.

In conclusion, our work demonstrates that reinforcement learning is a suitable, promising learning paradigm to study the emergence of collective phenomena in adaptive stochastic many-body systems. In Chapter 4 we discuss how this proof-of-concept work directly leads to exciting research possibilities and how they could be implemented in a practical setting.

4. Conclusions and future perspectives

La sabiduría nos llega cuando ya no nos sirve de nada.

Gabriel García Márquez - El amor en los tiempos del cólera

Natural phenomena span an enormous range of spatio-temporal scales, from the ultra-fast dynamics of the fundamental constituents of matter to the slow evolution of the structure of our universe. The dynamics of physical systems result from the feedback between processes taking place at these different levels of organisation. In this thesis we have studied the interplay between macroscopic and microscopic degrees of freedom in two different contexts, a biological and an artificial one, corresponding to the two chapters of this dissertation. In the following we summarize the main conclusions of this thesis and discuss possible directions of future research.

In the first part of this thesis (Chapter 2) we studied how a biological system, the primitively social wasp *Polistes canadensis* (Figure 2.1), can simultaneously achieve robust specialisation in the presence of noise and display strong phenotypic plasticity after queen removal using a combination of a multiscale experimental approach with a theoretical description of the system motivated by the experimental evidence. Remarkably, our unique experimental approach allowed us to correlate molecular, anatomical and behavioural measurements at the level of single insects (Figure 2.3).

In order to rigorously understand the mechanism underlying the regulation of specialization and plasticity in *Polistes*, we built a minimal theoretical description of the stochastic dynamics of the system using a master equation formalism. For the sake of simplicity, we did not assume any non-linear effects unless explicitly required by the empirical observations. The master equation formalism provides a description of the system dynamics that makes accurate predictions of the temporal evolution of experimental observables such as the number of egg layers over time or the distribution of ovary sizes across the population (Figure 2.16).

To understand the structure of the phase space and the response of the system to perturbations we derived a mean-field master equation in the continuum limit that retains the main features of the full model and represented the coevolution of the population and molecular degrees of freedom as a phase portrait (Figure 2.12). Our findings reveal that specialization

and plasticity are emergent properties resulting from antagonistic interactions at the population and the molecular scales. Additionally, we showed that the system is stable against intrinsic perturbations, that do not modify the population composition, while remaining plastic against extrinsic perturbations. In particular, the system undergoes a functional saddle node bifurcation where the role of control parameter is played by the population structure itself.

Finally, we studied how stochastic fluctuations can destabilise the social structure. Our calculations indicate that a small society regulated exclusively by stochastic interactions with a rate comparable to the time needed to upregulate the queen genes should not be stable over long periods of time. Therefore, we concluded that for the society to be stable over long periods of time additional layers of molecular regulation that stabilize individual gene expression profiles are needed. We hypothesized that DNA methylation, an epigenetic DNA modification, could stabilise the social structure by decreasing gene expression variability across the population and showed that this fact would lead to an exponential increase in the lifetime of the society (Figure 2.21).

Our approach highlights the crucial role of multiscale regulation in biological systems. In order to understand the dynamics of the system, the combination of integrative behavioural, physiological and molecular measurements and an interdisciplinary approach is required. In the following paragraphs we discuss future research directions inspired by this work.

Although our theoretical description relies on the interplay between processes at different spatial scales, it only accurately describes the dynamics of the system during the time scale associated with the reprogramming process. In fact, a description of the nest dynamics over longer time scales must take into account birth and death of insects. It has been empirically observed that over longer periods of time the number of insects in the nest grows due to the birth of new workers, leading to a state characterized by the presence of a large number of workers and a single queen. In this state, workers do not engage in subdominant interactions with the queen often enough, giving rise to the emergence of a second queen-like individual that destabilises the colony. In the last phase of this process the two queen-like individuals try to establish themselves as dominant, leading to the splitting of the nest into two groups. An interesting extension of this project would be to characterize this dynamical instability, extending our theoretical description to account for processes happening at longer time scales.

Additionally, in this project we have developed a mean-field theory, neglecting any spatial dependence of the interaction kernel. However, the role of spatial structures and local interactions is crucial in the regulation of other biological systems such as tissues. In collaboration

with an intern, Misha Gupta, we have started developing a spatial model where individuals diffuse in space and locally interact according to the same interaction kernel. Preliminary results indicate that such a system exhibits a patterned steady-state for suitable combinations of the parameters. Our aim is to uncover the connection between patterning and functional bifurcations in this context, as it has already been recently explored in other active matter systems [80].

Finally, the insights gained from this project are not limited to the study of social insects. Our approach also anticipates future experimental developments. Research on cellular aggregates, such as tissues or organoids, currently lacks technologies that simultaneously probe the dynamics at multiple levels of biological organisation, albeit the existence of such technologies, such as RNA sequencing or lineage tracing, at single scales [143–146]. Given the current interest in the development of multiomics technologies, we believe our approach can also prove useful in the study of the dynamics of such systems.

In the second part of the thesis (Chapter 3) we have studied how macroscopic feedback can remodel microscopic degrees of freedom in an artificial system. Additionally, we were interested in understanding how this remodelling could lead to the emergence of collective behaviour. In particular, to simulate the active remodelling of microscopic degrees of freedom in a dynamic environment we have developed an approach that combines stochastic dynamics with deep reinforcement learning. We consider a many-particle system where the reaction rates that define the stochastic dynamics of each particle are calculated using neural networks that are updated using reinforcement learning (Figure 3.3). For the sake of clarity, we have focused on a 1D lattice system where particles are trained to maximize a reward function that penalizes collisions.

Our results reveal the existence of a steady-state characterized by the presence of two groups of particles with positive and negative asymptotic velocities (Figure 3.13) whose formation relies on an active feedback, as indicated by the reduction of the variance of the number of particles in each of the two groups compared with a null model (Figure 3.14). The particle dynamics exhibit a cross-over between early diffusive and late ballistic regimes (Figure 3.16), that corresponds to the emergence of a non-zero average velocity.

Additionally, we showed that the average nearest neighbour distance decreases monotonically over time (Figure 3.17) as a result of the development of interactions that are attractive over long distances and become repulsive over short distances. The counter-intuitive steady-state dynamics can be rationalized as a result of stochastic fluctuations destabilising the deterministic optimal strategy.

The left/right symmetry breaking at the level of individual particles in the steady-state is also reflected at the microscopic level, represented by the structure of the neural networks that determine the individual reaction rates, as indicated by a low dimensional representation of the neural network parameters that shows the microscopic divergence of the neural networks of particles with positive and negative asymptotic velocity (Figure 3.19).

Finally, we explored the impact of memory and demographic disorder in the dynamics of the system. Remarkably, memory effects, resulting from a non-zero discount parameter, give rise to a transition between two different asymptotic dynamical regimes (Figure 3.23). Whereas for small values of the discount parameter the sign of the asymptotic velocity of individual particles is well defined, for larger values of the discount parameter the velocity of individual particles changes sign over time, with the transition happening for a value of the discount parameter of approximately $\gamma \approx 0.45$. The existence of the two regimes is reflected in the non-monotonic behaviour of the autocorrelation time of single velocities, that exhibits a local maximum at the transition point (Figure 3.25).

In order to study the effects of demographic disorder in the dynamics of the system we introduced a second species of collision-seeking particles. In this case, for small values of the discount parameter, the dynamics lead to the formation of structures composed by a dense core of collision-seeking particles surrounded by a dilute region of collision-avoiding particles that are stable over long times (Figure 3.26).

In conclusion, we leveraged recent developments in the field of artificial intelligence to establish a minimal model to study the remodelling of microscopic degrees of freedom as a result of macroscopic feedback that could prove useful for both the physics and the computer science communities. In the following paragraphs we discuss future research directions inspired by this work.

In this thesis we have focused on studying the dynamics of a one dimensional non-equilibrium system. Therefore, one possible extension of our work would be to study the dependence of the phenomenology that we observed with the dimensionality of the system. In two-dimensional equilibrium systems a central result of statistical mechanics, the Mermin-Wagner theorem [147], states that the spontaneous symmetry breaking of a continuous symmetry is not possible. However, in the non-equilibrium case continuous symmetries can be spontaneously broken giving rise to an ordered phase with a continuous order parameter [148–150]. In collaboration of Onurcan Bektas, we are implementing 2D stochastic simulations to explore whether reinforcement learning and collision avoidance can lead to the emergence of an ordered phase.

Although here we have focused on a system with explicit spatial dependence, our approach does not rely on it to induce interactions, allowing us to realistically model systems with a large number of microscopic degrees of freedom. Machine learning provides a practical way of representing large number of degrees of freedom without neglecting their intrinsic complexity that can be used to model a wealth of physical systems by adequately defining the reward function. Our approach based on reinforcement learning also has the appeal of simplicity, as rich dynamics can result from simple reward functions.

Our results reveal the fundamental role played by macroscopic feedback in the stabilization of microscopic states both in nature and *in silico*. We hope that the new possibilities offered by our results at both the experimental and theoretical levels will keep sparking interest in these fascinating systems for a long time.

Bibliography

1. Kolmogorov, A. N. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Cr Acad. Sci. URSS* **30**, 301–305 (1941).
2. Leonard, A. Energy cascade in large-eddy simulations of turbulent fluid flows. **18**, 237–248 (1975).
3. Paret, J. & Tabeling, P. Experimental observation of the two-dimensional inverse energy cascade. *Physical review letters* **79**, 4162 (1997).
4. Kadanoff, L. P. Scaling Laws for Ising Models near T c. *Physics Physique Fizika* **2**, 263–272 (1966).
5. Wilson, K. G. Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Physical Review B* **4**, 3174–3183 (1971).
6. Wilson, K. G. The Renormalization Group and Critical Phenomena. *Reviews of Modern Physics* **55**, 583–600 (1983).
7. Goldenfeld, N. *Lectures on Phase Transitions and the Renormalization Group Frontiers in Physics* v. **85** (Addison-Wesley, Advanced Book Program, Reading, Mass, 1992).
8. Kardar, M. *Statistical Physics of Fields* ISBN: 978-0-521-87341-3 (Cambridge University Press, Cambridge ; New York, 2007).
9. Täuber, U. C. *Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior* ISBN: 978-0-521-84223-5 (Cambridge University Press, Cambridge, United Kingdom ; New York, 2014).
10. Ramaswamy, S. The Mechanics and Statistics of Active Matter. *Annual Review of Condensed Matter Physics* **1**, 323–345 (2010).
11. Marchetti, M. C. *et al.* Hydrodynamics of Soft Active Matter. *Reviews of Modern Physics* **85**, 1143–1189 (2013).
12. Fodor, É. *et al.* How Far from Equilibrium Is Active Matter? *Physical Review Letters* **117**, 038103 (2016).

13. Schrodinger, R., Schrödinger, E. & Dinger, E. S. *What is life?: With mind and matter and autobiographical sketches* (Cambridge university press, 1992).
14. Jopling, C., Boue, S. & Belmonte, J. C. I. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nature reviews Molecular cell biology* **12**, 79–89 (2011).
15. Tata, P. R. *et al.* Dedifferentiation of Committed Epithelial Cells into Stem Cells in Vivo. *Nature* **503**, 218–223 (2013).
16. Lang, A. H., Li, H., Collins, J. J. & Mehta, P. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLoS computational biology* **10**, e1003734 (2014).
17. Hänggi, P., Talkner, P. & Borkovec, M. Reaction-Rate Theory: Fifty Years after Kramers. *Reviews of Modern Physics* **62**, 251–341 (1990).
18. Cross, M. & Greenside, H. *Pattern Formation and Dynamics in Nonequilibrium Systems* Cambridge (2009).
19. Assaf, M. & Meerson, B. WKB Theory of Large Deviations in Stochastic Populations. *Journal of Physics A: Mathematical and Theoretical* **50**, 263001 (2017).
20. West-Eberhard, M. J. Phenotypic plasticity and the origins of diversity. *Annual review of Ecology and Systematics* **20**, 249–278 (1989).
21. DeWitt, T. J., Sih, A. & Wilson, D. S. Costs and Limits of Phenotypic Plasticity. *Trends in Ecology & Evolution* **13**, 77–81 (1998).
22. Schlichting, C. D. The evolution of phenotypic plasticity in plants. *Annual review of ecology and systematics* **17**, 667–693 (1986).
23. Oliveira, R. F. Social Plasticity in Fish: Integrating Mechanisms and Function. *Journal of Fish Biology* **81**, 2127–2150 (2012).
24. West-Eberhard, M. J. *Developmental Plasticity and Evolution* Oxford Uni (2003).
25. Beldade, P., Mateus, A. R. A. & Keller, R. A. Evolution and Molecular Mechanisms of Adaptive Developmental Plasticity. *Molecular Ecology* **20**, 1347–1363 (2011).
26. Cardoso, S. D., Teles, M. C. & Oliveira, R. F. Neurogenomic mechanisms of social plasticity. *Journal of Experimental Biology* **218**, 140–149 (2015).
27. von Frisch, K. Decoding the Language of the Bee. *Science* **185**, 663–668 (1974).
28. Grueter, C. & Leadbeater, E. Insights from insects about adaptive social information use. *Trends in ecology & evolution* **29**, 177–184 (2014).

29. Gardner, K. E., Seeley, T. D. & Calderone, N. W. Do honeybees have two discrete dances to advertise food sources? *Animal Behaviour* **75**, 1291–1300 (2008).
30. Tennenbaum, M., Liu, Z., Hu, D. & Fernandez-Nieves, A. Mechanics of Fire Ant Aggregations. *Nature Materials* **15**, 54–59 (2016).
31. Peleg, O., Peters, J. M., Salcedo, M. K. & Mahadevan, L. Collective Mechanical Adaptation of Honeybee Swarms. *Nature Physics* **14**, 1193–1198 (2018).
32. Buhl, J. *et al.* From Disorder to Order in Marching Locusts. *Science* **312**, 1402–1406 (2006).
33. Mlot, N. J., Tovey, C. A. & Hu, D. L. Fire Ants Self-Assemble into Waterproof Rafts to Survive Floods. *Proceedings of the National Academy of Sciences* **108**, 7669–7673 (2011).
34. Cronin, A. L. Social Flexibility in a Primitively Social Allodapine Bee (Hymenoptera: Apidae): Results of a Translocation Experiment. *Oikos* **94**, 337–343 (2001).
35. Sumner, S., Pereboom, J. J. & Jordan, W. C. Differential Gene Expression and Phenotypic Plasticity in Behavioural Castes of the Primitively Eusocial Wasp, *Polistes Canadensis*. *Proc Biol Sci* **273**, 19–26 (2006).
36. Corona, M., Libbrecht, R. & Wheeler, D. E. Molecular Mechanisms of Phenotypic Plasticity in Social Insects. *Current Opinion in Insect Science* **13**, 55–60 (2016).
37. Patalano, S. *et al.* Molecular Signatures of Plastic Phenotypes in Two Eusocial Insect Species with Simple Societies. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13970–13975 (2015).
38. Suzuki, R. & Arita, T. A simple computational model of the evolution of a communicative trait and its phenotypic plasticity. *Journal of theoretical biology* **330**, 37–44 (2013).
39. Xue, B. & Leibler, S. Benefits of Phenotypic Plasticity for Population Growth in Varying Environments. *Proceedings of the National Academy of Sciences* **115**, 12745–12750 (2018).
40. Xue, B., Sartori, P. & Leibler, S. Environment-to-Phenotype Mapping and Adaptation Strategies in Varying Environments. *Proceedings of the National Academy of Sciences* **116**, 13847–13855 (2019).
41. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* ISBN: 978-0-262-03561-3 (The MIT Press, Cambridge, Massachusetts, 2016).

42. Carleo, G. *et al.* Machine Learning and the Physical Sciences. *Reviews of Modern Physics* **91**, 045002 (2019).
43. Mehta, P. *et al.* A High-Bias, Low-Variance Introduction to Machine Learning for Physicists. *Physics Reports* **810**, 1–124 (2019).
44. Pang, B., Lee, L. & Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. *arXiv:cs/0205070*. arXiv: [cs/0205070](https://arxiv.org/abs/cs/0205070) (2002).
45. Kotsiantis, S. B., Zaharakis, I., Pintelas, P., *et al.* Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* **160**, 3–24 (2007).
46. Ranzato, M., Huang, F. J., Boureau, Y.-L. & LeCun, Y. *Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition in 2007 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Minneapolis, MN, USA, 2007), 1–8.
47. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).
48. Kirkpatrick, J. *et al.* Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences* **114**, 3521–3526 (2017).
49. Shore, J. & Johnson, R. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Transactions on Information Theory* **26**, 26–37 (1980).
50. Hertz, J., Krogh, A. & Palmer, R. G. *Introduction to the Theory of Neural Computation* First (CRC Press, 2018).
51. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* Second edition (The MIT Press, Cambridge, Massachusetts, 2018).
52. Mnih, V. *et al.* Human-Level Control through Deep Reinforcement Learning. *Nature* **518**, 529–533 (2015).
53. Silver, D. *et al.* A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play. *Science* **362**, 1140–1144 (2018).
54. Ocko, S. A., Heyde, A. & Mahadevan, L. Morphogenesis of Termite Mounds. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 3379–3384 (2019).

-
55. Merrell, A. J. & Stanger, B. Z. Adult Cell Plasticity in Vivo: De-Differentiation and Transdifferentiation Are Back in Style. *Nature Reviews Molecular Cell Biology* **17**, 413–425 (2016).
 56. Kennedy, P. *et al.* Deconstructing Superorganisms and Societies to Address Big Questions in Biology. *Trends in Ecology & Evolution* **32**, 861–872 (2017).
 57. Herb, B. *et al.* Reversible Switching between Epigenetic States in Honeybee Behavioral Subcastes. *Nature neuroscience* **15**, 1371–1373 (2012).
 58. Simola, D. F. *et al.* Epigenetic (Re)Programming of Caste-Specific Behavior in the Ant *Camponotus Floridanus*. *Science* **351** (2016).
 59. Todd, E. V. *et al.* Stress, Novel Sex Genes, and Epigenetic Reprogramming Orchestrate Socially Controlled Sex Change. *Science Advances* **5** (2019).
 60. West-Eberhard, M. J. *The Social Biology of Polistine Wasps* Mich., Mis (ed C. Evans, F.) (1969).
 61. Turillazzi, S. & West-Eberhard, M. J. *Natural History and Evolution of Paper Wasps* (Oxford, 1996).
 62. Patalano, S., Hore, T. A., Reik, W. & Sumner, S. Shifting Behaviour: Epigenetic Reprogramming in Eusocial Insects. *Curr Opin Cell Biol* **24**, 367–373 (2012).
 63. Ferreira, P. G. *et al.* Transcriptome Analyses of Primitively Eusocial Wasps Reveal Novel Insights into the Evolution of Sociality and the Origin of Alternative Phenotypes. *Genome biology* **14**, R20 (2013).
 64. Patalano, S. *et al.* Specialisation and Plasticity in a Primitively Social Insect (2020).
 65. Taylor, B. A., Cini, A., Wyatt, C. D. R., Reuter, M. & Sumner, S. The Molecular Basis of Socially Mediated Phenotypic Plasticity in a Eusocial Paper Wasp. *Nature Communications* **12**, 775 (2021).
 66. Hunt, J. H. *et al.* Differential Gene Expression and Protein Abundance Evince Ontogenetic Bias toward Castes in a Primitively Eusocial Wasp. *PLoS One* **5** (2010).
 67. Holman, L., Helanterä, H., Trontti, K. & Mikheyev, A. S. Comparative Transcriptomics of Social Insect Queen Pheromones. *Nature Communications* **10**, 1593 (2019).
 68. Hamilton, A. R., Shpigler, H., Bloch, G., Wheeler, D. E. & Robinson, G. E. *Endocrine Influences on Insect Societies* 421–451 (2016).
 69. Theraulaz, G., Pratte, M. & Gervet, J. Behavioural Profiles in *Polistes Dominulus* (Christ) Wasp Societies: A Quantitative Study. *Behaviour* **113** (1990).

70. Strassmann, J. E. *et al.* The Cost of Queen Loss in the Social Wasp *Polistes Dominulus* (Hymenoptera: Vespidae). *Journal of the Kansas entomological society* **77**, 343–355 (2004).
71. Roeseler, P.-F., Roeseler, I., Strambi, A. & Augier, R. Influence of Insect Hormones on the Establishment of Dominance Hierarchies among Foundresses of the Paper Wasp, *Polistes Gallicus*. *Behav Ecol Sociobiol* **15**, 133–142 (1984).
72. Giray, T., Giovanetti, M. & West-Eberhard, M. J. Juvenile Hormone, Reproduction, and Worker Behavior in the Neotropical Social Wasp *Polistes Canadensis*. *Proceedings of the National Academy of Sciences* **102**, 3330–3335 (2005).
73. Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry* **81**, 2340–2361 (1977).
74. Démery, V., Bénichou, O. & Jacquin, H. Generalized Langevin Equations for a Driven Tracer in Dense Soft Colloids: Construction and Applications. *New Journal of Physics* **16**, 053032 (2014).
75. Poncet, A., Bénichou, O., Démery, V. & Oshanin, G. Bath-Mediated Interactions between Driven Tracers in Dense Single Files. *Physical Review Research* **1**, 033089 (2019).
76. Gardiner, C. W. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer-Verlag, 2004).
77. Bauer, M., Knebel, J., Lechner, M., Pickl, P. & Frey, E. Ecological Feedback in Quorum-Sensing Microbial Populations Can Induce Heterogeneous Production of Autoinducers. *eLife* **6**, 1–38 (2017).
78. Jörg, D. J., Kitadate, Y., Yoshida, S. & Simons, B. D. Competition for Stem Cell Fate Determinants as a Mechanism for Tissue Homeostasis. *arXiv:1901.03903 [physics, q-bio]*. arXiv: 1901.03903 [physics, q-bio] (2019).
79. Rulands, S., Klünder, B. & Frey, E. Stability of Localized Wave Fronts in Bistable Systems. *Physical Review Letters* **110**, 038102 (2013).
80. Denk, J. & Frey, E. Pattern-Induced Local Symmetry Breaking in Active Matter Systems. *arXiv:2005.12791 [cond-mat, physics:physics]*. arXiv: 2005.12791 [cond-mat, physics:physics] (2020).
81. Southon, R. J., Bell, E. F., Graystock, P. & Sumner, S. Long Live the Wasp: Adult Longevity in Captive Colonies of the Eusocial Paper Wasp *Polistes Canadensis* (L.). *PeerJ* **3**, e848 (2015).

-
82. Alberts, B. *Molecular Biology of the Cell 6th Edition* (2015).
 83. Bird, A. The Essentials of DNA Methylation. *Cell* **70**, 5–8 (1992).
 84. Smallwood, S. A. *et al.* Single-Cell Genome-Wide Bisulfite Sequencing for Assessing Epigenetic Heterogeneity. *Nature Methods* **11**, 817–820 (2014).
 85. Reik, W. Epigenetic Reprogramming in Mammalian Development. *Science* **293**, 1089–1093 (2001).
 86. Parry, A., Rulands, S. & Reik, W. Active Turnover of DNA Methylation during Cell Fate Decisions. *Nature Reviews Genetics* **22**, 59–66 (2021).
 87. Smith, Z. D. & Meissner, A. DNA Methylation: Roles in Mammalian Development. *Nature Reviews Genetics* **14**, 204–220 (2013).
 88. Das, P. M. & Singal, R. DNA Methylation and Cancer. *Journal of Clinical Oncology* **22**, 4632–4642 (2004).
 89. Robertson, K. D. DNA Methylation and Human Disease. *Nature Reviews Genetics* **6**, 597–610 (2005).
 90. Huh, I., Zeng, J., Park, T. & Yi, S. V. DNA Methylation and Transcriptional Noise. *Epigenetics & Chromatin* **6**, 9 (2013).
 91. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128**, 707–719 (2007).
 92. Glastad, K. M. *et al.* Variation in DNA Methylation Is Not Consistently Reflected by Sociality in Hymenoptera. *Genome Biology and Evolution* **9**, 1687–1698 (2017).
 93. Biró, T. & Néda, Z. Unidirectional Random Growth with Resetting. *Physica A: Statistical Mechanics and its Applications* **499**, 335–361 (2018).
 94. Tindo, M. & Dejean, A. Dominance Hierarchy in Colonies of *Belonogaster Juncea Juncea* (Vespidae, Polistinae). *Insectes Sociaux* **47**, 158–163 (2000).
 95. Halliwell, S. & Aristotle, A. *Aristotle's poetics* (University of Chicago Press, 1998).
 96. Bacon, F. *The advancement of learning [1605]* (Harvard University Press, 2013).
 97. Skinner, B. F. *Science and Human Behavior* (The Free Press, New York, 1965).
 98. Cooper, J. O., Heron, T. E. & Heward, W. L. *Applied Behavior Analysis* Third edition. ISBN: 978-0-13-475255-6 (Pearson, Hoboken, New Jersey, 2019).
 99. Watson, J. B. & Rayner, R. Conditioned Emotional Reactions. *Journal of Experimental Psychology* **3**, 1–14 (1920).

100. Skinner, B. F. 'Superstition' in the Pigeon. *Journal of Experimental Psychology* **38**, 168–172 (1948).
101. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* **65**, 386–408 (1958).
102. Silver, D., Singh, S., Precup, D. & Sutton, R. S. Reward Is Enough. *Artificial Intelligence* **299**, 103535 (2021).
103. Mnih, V. *et al.* Playing Atari with Deep Reinforcement Learning. arXiv: 1312.5602 [cs] (2013).
104. Mnih, V. *et al.* Asynchronous Methods for Deep Reinforcement Learning. arXiv: 1602.01783 [cs] (2016).
105. Kober, J., Bagnell, J. A. & Peters, J. Reinforcement Learning in Robotics: A Survey. *The International Journal of Robotics Research* **32**, 1238–1274 (2013).
106. Sadler, M. & Regan, N. *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI* (New in Chess, 2019).
107. Nir, A., Sela, E., Beck, R. & Bar-Sinai, Y. Machine-Learning Iterative Calculation of Entropy for Physical Systems. arXiv: 2008.05921 [cond-mat] (2020).
108. Seif, A., Hafezi, M. & Jarzynski, C. Machine Learning the Thermodynamic Arrow of Time. *Nature Physics* **17**, 105–113 (2021).
109. Gordon, A., Banerjee, A., Koch-Janusz, M. & Ringel, Z. Relevance in the Renormalization Group and in Information Theory. arXiv:2012.01447 [cond-mat]. arXiv: 2012.01447 [cond-mat] (2020).
110. Lenggenhager, P. M., Gökmen, D. E., Ringel, Z., Huber, S. D. & Koch-Janusz, M. Optimal Renormalization Group Transformation from Information Theory. *Physical Review X* **10**, 011037 (2020).
111. Gökmen, D. E., Ringel, Z., Huber, S. D. & Koch-Janusz, M. Statistical Physics through the Lens of Real-Space Mutual Information. arXiv:2101.11633 [cond-mat]. arXiv: 2101.11633 [cond-mat] (2021).
112. Gabrié, M. *et al.* Entropy and Mutual Information in Models of Deep Neural Networks. *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 124014 (2019).
113. Goldt, S., Advani, M. S., Saxe, A. M., Krzakala, F. & Zdeborová, L. Dynamics of Stochastic Gradient Descent for Two-Layer Neural Networks in the Teacher–Student Setup. *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 124010 (2020).

-
114. Feng, Y. & Tu, Y. The Inverse Variance–Flatness Relation in Stochastic Gradient Descent Is Critical for Finding Flat Minima. *Proceedings of the National Academy of Sciences* **118**, e2015617118 (2021).
 115. Pezzotta, A., Adorisio, M. & Celani, A. Chemotaxis Emerges as the Optimal Solution to Cooperative Search Games. *Physical Review E* **98**, 042401 (2018).
 116. Reddy, G., Wong-Ng, J., Celani, A., Sejnowski, T. J. & Vergassola, M. Glider Soaring via Reinforcement Learning in the Field. *Nature* **562**, 236–239 (2018).
 117. Mazzolini, A. & Celani, A. Generosity, Selfishness and Exploitation as Optimal Greedy Strategies for Resource Sharing. *arXiv:1903.02786 [physics, q-bio]*. arXiv: 1903.02786 [physics, q-bio] (2019).
 118. Durve, M., Peruani, F. & Celani, A. Learning to Flock through Reinforcement. *Physical Review E* **102**, 012601 (2020).
 119. Senior, A. W. *et al.* Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **577**, 706–710 (2020).
 120. Jumper, J. *et al.* Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* (2021).
 121. Tunyasuvunakool, K. *et al.* Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* (2021).
 122. Bellman, R. Dynamic programming. *Science* **153**, 34–37 (1966).
 123. Watkins, C. J. C. H. *Learning from Delayed Rewards* PhD thesis (University of Cambridge, 1989).
 124. Watkins, C. J. C. H. & Dayan, P. Q-Learning. *Machine Learning* **8**, 279–292 (1992).
 125. Rummery, G. & Niranjan, M. *On-Line Q-Learning Using Connectionist Systems* Technical Report CUED/F-INFENG/TR 166 (University of Cambridge, Engineering Department, 1994).
 126. Sutton, R. S. *Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding in Advances in Neural Information Processing Systems 8* (MIT Press, 1996), 1038–1044.
 127. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* **86**, 2278–2324 (1998).
 128. Bengio, Y., Ducharme, R. & Vincent, P. *A neural probabilistic language model in Advances in Neural Information Processing Systems* (2001), 932–938.

129. Min, S., Lee, B. & Yoon, S. Deep Learning in Bioinformatics. *Briefings in Bioinformatics* (2016).
130. Ji, Y., Lotfollahi, M., Wolf, F. A. & Theis, F. J. Machine Learning for Perturbational Single-Cell Omics. *Cell Systems* **12**, 522–537 (2021).
131. Lotfollahi, M. *et al.* Mapping Single-Cell Data to Reference Atlases by Transfer Learning. *Nature Biotechnology* (2021).
132. Meier, F. *et al.* Deep Learning the Collisional Cross Sections of the Peptide Universe from a Million Experimental Values. *Nature Communications* **12**, 1185 (2021).
133. LeCun, Y. *Generalization and Network Design Strategies* tech. rep. CRG-TR-89-4 (University of Toronto, University of Toronto, 1989).
134. Schmidhuber, J. Learning Complex, Extended Sequences Using the Principle of History Compression. *Neural Computation* **4**, 234–242 (1992).
135. Cybenko, G. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems* **2**, 303–314 (1989).
136. Pinkus, A. Approximation Theory of the MLP Model in Neural Networks. *Acta Numerica* **8**, 143–195 (1999).
137. Lu, Z., Pu, H., Wang, F., Hu, Z. & Wang, L. *The Expressive Power of Neural Networks: A View from the Width* in *Advances in Neural Information Processing Systems* **30** (Curran Associates, Inc., 2017).
138. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **323**, 533–536 (1986).
139. Bottou, L. in *Neural Networks: Tricks of the Trade* 421–436 (Springer, 2012).
140. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. arXiv: 1412.6980 [cs] (2017).
141. Chen, T. *et al.* MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. *arXiv:1512.01274 [cs]*. arXiv: 1512.01274 [cs] (2015).
142. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*. arXiv: 1802.03426 [cs, stat] (2020).
143. Blanpain, C. & Simons, B. D. Unravelling Stem Cell Dynamics by Lineage Tracing. *Nature Reviews Molecular Cell Biology* **14**, 489–502 (2013).

-
144. Kester, L. & van Oudenaarden, A. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* **23**, 166–179 (2018).
 145. Kretzschmar, K. & Watt, F. M. Lineage Tracing. *Cell* **148**, 33–45 (2012).
 146. Wagner, D. E. & Klein, A. M. Lineage Tracing Meets Single-Cell Omics: Opportunities and Challenges. *Nature Reviews Genetics* **21**, 410–427 (2020).
 147. Mermin, N. D. & Wagner, H. Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models. *Physical Review Letters* **17**, 1133–1136 (1966).
 148. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. Novel Type of Phase Transition in a System of Self-Driven Particles. *Physical Review Letters* **75**, 1226–1229 (1995).
 149. Toner, J. & Tu, Y. Long-Range Order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together. *Physical Review Letters* **75**, 4326–4329 (1995).
 150. Toner, J. & Tu, Y. Flocks, Herds, and Schools: A Quantitative Theory of Flocking. *Physical Review E* **58**, 4828–4858 (1998).

A. Numerical simulations of the master equations

In order to obtain approximate solutions of the master equations of Chapter 2 for a population of N individuals, we performed kinetic Monte Carlo simulations following Gillespie's algorithm [73]. We implemented this algorithm in Julia using custom code. Out of the different possible reactions - production of a new gene product, degradation of gene products, interactions, degradation of the queen gene repressors or ovary development - one was randomly selected with probability proportional to the overall rate of the respective process. These rates can be written in vector form as

$$R = \left(\mu, \delta n_i, \alpha K(n_i, n_j), \delta(t_{int}^i - t_{per}), g(n_i) \right) \quad (\text{A.1})$$

After a process has been selected the simulation time t is increased by an amount Δt drawn from an exponential distribution of parameter λ given by the inverse sum of the overall rates

$$\lambda^{-1} = \sum_i \left(\mu + \delta n_i + \sum_j K(n_i, n_j) + \delta(t_{int}^i - t_{per}) + g(n_i) \right) \quad (\text{A.2})$$

Finally, the state of the system is updated based on the selected reaction and the process is repeated until the maximum allotted time t_{\max} has been exceeded.

The following parameters were used for all the simulations $\mu = 500$, $\delta = 1$, $n_0 = 250$, $\lambda = 10$, $t_{per} = 1$ unless otherwise specified. Additionally, the ovary growth rate was chosen so that ovaries were mature in 6 days, as observed experimentally.

Below we provide a pseudocode of our implementation of the Gillespie stochastic simulation algorithm

Algorithm 1 Gillespie algorithm

- 1: **while** $t < t_{\max}$ **do**
 - 2: Calculate rates $R = (\mu, \delta n_i, \alpha K(n_i, n_j), \delta(t_{int}^i - t_{per}), g(n_i))$
 - 3: $\lambda^{-1} = \sum_i (\mu + \delta n_i + \sum_j K(n_i, n_j) + \delta(t_{int}^i - t_{per}) + g(n_i))$
 - 4: $\Delta t \sim \text{Exp}(\lambda)$
 - 5: $t \rightarrow t + \Delta t$
 - 6: Select reaction r_k with probabilities given by R
 - 7: Update (n_i, o_i) according to r_k
-

Versicherung

Hiermit versichere ich, Adolfo Alsina López, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Diese Arbeit wurde unter der wissenschaftlichen Betreuung von Dr. Steffen Rulands am Max-Planck Institut für Physik komplexer Systeme in Dresden angefertigt. Ich erkläre hiermit, dass keine früheren erfolgreichen Promotionsverfahren stattgefunden haben. Ich erkenne die Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften der Technische Universität Dresden an.

Unterschrift

Dresden, Oktober 2021