UNIVERSIDADE FEDERAL DE PERNAMBUCO
DEPARTAMENTO DE FÍSICA – CCEN
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

ANDRÉ DA CONCEIÇÃO AMADO

FITNESS TRADEOFFS IN THE EVOLUTIONARY TRANSITION TO MULTI-CELLULARITY AND THE EVOLUTION OF COMPLEXITY

Recife
2018

ANDRÉ DA CONCEIÇÃO AMADO

FITNESS TRADEOFFS IN THE EVOLUTIONARY TRANSITION TO MULTICELLULARITY AND THE EVOLUTION OF COMPLEXITY

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Física.

Orientador: Prof. Dr. Paulo Roberto de Araujo Campos

Recife
2018

Catalogação na fonte
Bibliotecária Elaine Cristina de Freitas CRB4-1790

A481f Amado, André da Conceição

Fitness tradeoffs in the evolutionary transition to multicellularity and the evolution of complexity / André da Conceição Amado. – 2018.

138 f.

Orientador: Paulo Roberto de Araújo Campos

Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN. Física. Recife, 2018.

Inclui referências e apêndice

1. Física estatística 2. Dinâmica evolucionária 3. Tradeoffs 4. Multicelularidade I. Campos, Paulo Roberto de Araújo (Orientador) II. Título.

530.1595 CDD (22. ed.) UFPE-FQ 2018-23

ANDRÉ DA CONCEIÇÃO AMADO

FITNESS TRADEOFFS IN THE EVOLUTIONARY TRANSITION TO MULTICELLULARITY AND THE EVOLUTION OF COMPLEXITY

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Física.

Aprovada em: 23/02/2018.

BANCA EXAMINADORA

____
Prof. Dr. Paulo Roberto de Araujo Campos
Orientador
Universidade Federal de Pernambuco

____
Prof. Dr. Ernesto Carneiro Pessoa Raposo
Examinador Interno
Universidade Federal de Pernambuco

____
Prof. Dr. Mauro Copelli Lopes da Silva
Examinador Interno
Universidade Federal de Pernambuco

____
Prof. Dr. Jeferson Jacob Arenzon
Examinador Externo
Universidade Federal do Rio Grande do Sul

____
Prof. Dr. Marcus Aloizio Martinez de Aguiar
Examinador Externo
Universidade Estadual de Campinas

À minha avó,

À Diana, ser vivo que, estranhamente, não é mais uma menininha.

AGRADECIMENTOS

Embora tenha escrito a tese em Inglês pareceu-me mais natural expressar os meus agradecimentos na minha língua nativa. Um doutoramento é um processo impossível de completar sozinho. Felizmente, com todo o suporte e carinho que tive a meu lado foi um processo que decorreu de forma tranquila, sem grandes sobressaltos.

Queria começar por agradecer ao meu orientador, Paulo Campos. Aprendi muito com ele durante estes anos. Estou-lhe agradecido pelo apoio dado e pela atenção despendida com vista ao sucesso dos seus alunos. Em paralelo com a nossa relação de trabalho, por vezes exigente, ele foi paciente e compreensivo sempre que necessitei. Além da relação de trabalho, acho que se desenvolveu entre nós amizade que espero que perdure após este doutoramento.

Queria agradecer a contribuição dos restantes coautores dos trabalhos que publicámos durante este período. Em particular, queria agradecer ao meu colega de trabalho mais próximo Lenin.

Gostaria de agradecer ao Departamento de Física da UFPE as condições de trabalho e aos meus professores em particular o apoio disponibilizado e a dedicação à qualidade das aulas.

Aos meus colegas e amigos no departamento queria agradecer o espírito de entreajuda e partilha que existiu. Seja nos cafés partilhados nos corredores ou nas noites a estudar para os EGDs houve sempre um espaço para o humor. Partilhei muitos momentos bons com amigos, que vou guardar na memória com carinho. Um agradecimento especial aos amigos com quem partilhei casa e momentos de sã loucura.

Destino um grande obrigado à minha família, de quem tenho estado longe ao longo destes anos, pelo se apoio e carinho constantes. Vou-lhes sempre estar reconhecido pelo ambiente equilibrado e estimulante em que cresci, rodeado de carinho. Este ambiente foi determinante na curiosidade constante que desenvolvi desde cedo. Além disso, foram eles que me ensinaram a procurar ser sempre uma pessoa correta na minha relação com os outros. O meu único remorso ao escrever esta tese em Inglês é a impossibilidade que daí advém de a partilhar decentemente com a minha família.

To my incredible partner Azadeh I have an infinite number of reasons to thank for. For a start, her joyful laughter that spreads happiness around and helped me to digest any problem that might have showed up. Her reasonable voice, love and care were essential contributions to keep me going steadily through my PhD. Discussions with her are always interesting and fruitful.

Por fim, queria agradecer ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) por providenciar a bolsa sem a qual este doutoramento não teria sido possível.

André

ABSTRACT

Tradeoffs are one of the essential ingredients that shape the diversity of life on Earth. They are thought to create ecological niches and restrict the accessible evolutionary pathways in a nontrivial way. Nevertheless, much is still unknown about the way tradeoffs steer the course of evolution. Recent studies open a new venue to the empirical exploration of this subject. The access to the genetic content of life has been revolutionizing the knowledge across the whole Biology, bringing some answers and raising a lot of new questions. The whole concept of multicellular life has been extending from the beginning of the 1990's with the recognition of multicellular bacteria and numerous behaviors in the now shadowy region between unicellular and multicellular life. The approach of experimental evolution recently provided the first experimental insights into the process of transition from unicellular to multicellular life, by evolving multicellular organisms under controlled conditions in the laboratory. This current work aims to provide a contribution to the theoretical understanding of the role of tradeoffs in the transition to multicellularity and complexity development. For that, we introduce and explore some models tailored to elucidate some of the aspects of these transitions. Each of those models is explored through a combination of analytical and simulational methods, which allows us to extract further information. A first approach deals with the establishment of an efficient mode of metabolism within the context of competition with a rapid and inefficient mode. Usually, high rate inefficient metabolisms tend to dominate, therefore extra mechanisms are necessary to counteract this. Within a resource-based formulation, we study the effect of group structure in the population and find that with groups the efficient mode outcompetes the inefficient one in a broad domain of the parameter space. In the sequel, we analyze the contribution of tradeoffs to the evolution of complexity. It is empirically known that complex networks of tradeoffs are established at the cellular and metabolic level. In this context, a system with an arbitrary number of tradeoffs over a given number of tasks is investigated. We carry out a statistical analysis over different sets of parameters in order to examine the dependence of cell specialization on the number and strength of the tradeoffs. A concrete application of the model to the carbon-nitrogen fixation tradeoff in cyanobacteria is provided. At last, we introduce a mechanistic model for the dynamics of multicellular aggregates. We consider the existence of different microscopic mechanisms shaping multicellular aggregates. Particularly, the model is applied to the study of the size-complexity rule and interesting results follows from that approach. Depending on the geometry of the aggregates the size-complexity rule can be followed or not. We found that more fragile aggregates violate the rule and more robust ones obey it. Each of the works addressed here provides some answers and raises new issues to be explored in the future. For instance, what is the effect of germ-soma tradeoffs for the outcomes predicted in our models? Or, if the size-complexity rule can be violated under some circumstances, there exist additional mechanisms that can also have the same effect? These and other questions are raised and briefly discussed in the conclusions.

Keywords: Evolutionary dynamics. Tradeoffs. Multicellularity. Biological complexity.

RESUMO

Os *tradeoffs*[1] são um dos ingredientes essenciais na definição da biodiversidade na Terra. Pensa-se que desempenham um papel crucial na criação de nichos ecológicos e que restringem o espaço evolucionário acessível de formas complexas. No entanto, muito é ainda desconhecido sobre o modo como os *tradeoffs* guiam o curso da evolução. Estudos recentes abrem novas perspectivas empíricas sobre este assunto. O acesso ao conteúdo genético da vida tem revolucionado o conhecimento em todas as áreas da Biologia, introduzindo algumas respostas e um sem-número de novas questões. O conceito de vida multicelular tem sido extendido desde o início dos anos 1990, com o reconhecimento de bactérias multicelulares e variados comportamentos ao longo do espectro que se abre entre vida unicelular e multicelular. A abordagem da evolução experimental providenciou recentemente os primeiros experimentos onde a transição de unicelular para multicelular pode ser observada directamente, sob condições controladas em laboratório. O trabalho aqui apresentado tem como objectivo contribuir para a compreensão teórica do papel dos *tradeoffs* na transição para a multicelularidade e desenvolvimento da complexidade. Para tal, nós introduzimos e exploramos alguns modelos desenhados para elucidar alguns aspectos destas transições. Os modelos são explorados utilizando uma combinação de métodos analíticos e simulações numéricas, o que nos permite obter mais informação dos modelos. Uma primeira abordagem lida com o estabelecimento de um modo eficiente de metabolismo no contexto de competição com um modo rápido e ineficiente. Geralmente metabolismos rápidos e ineficientes tendem a dominar, sendo necessária a introdução de outros mecanismos para o contrariar. No contexto de uma formulação baseada em recursos, nós estudamos o efeito da estruturação da população em grupos e obtemos agora que o metabolismo eficiente passa a dominar numa grande região do espaço de parâmetros. Em seguida, analisamos a contribuição dos *tradeoffs* para a evolução da complexidade. Empiricamente, sabe-se que redes complexas de *tradeoffs* são estabelecidas a nível celular e metabólico. Neste contexto, investigamos um sistema com um número arbitrário de *tradeoffs* incidentes sobre um dado número de tarefas. Realizamos uma análise estatística sobre diferentes conjuntos de parâmetros com o objectivo de examinar a forma como a especialização celular depende do número e intensidade dos *tradeoffs*. Apresentamos ainda uma aplicação concreta do modelo ao *tradeoff* existente entre os processos de fixação de carbono e nitrogênio nas cianobactérias. For fim, introduzimos um modelo mecanístico para a dinâmica dos agregados multicelulares. Consideramos a existência de diferentes mecanismos microscópicos que controlam a evolução dos agregados multicelulares. Em particular, aplicamos o modelo ao estudo da regra do tamanho-complexidade e obtemos resultados interessantes. Dependendo da geometria considerada, a regra do tamanho-complexidade pode ser respeitada ou não. Cada um dos trabalhos desenvolvidos respondem algumas questões e levantam outras a explorar no futuro. Por exemplo, qual é o efeito do *tradeoff* entre funções reprodutivas e somáticas nos resultados obtidos? Ou, se a regra do tamanho-complexidade pode ser violada sob certas condições, existirão mecanismos adicionais que produzem o mesmo efeito? Estas e outras questões são levantadas e discutidas brevemente nas conclusões.

Palavras-chave: Dinâmica evolucionária. Tradeoffs. Multicelularidade. Complexidade biológica.

[1] Optámos por manter a palavra *tradeoff* em Inglês mesmo na versão em Português devido à dificuldade de expressar o conceito de forma concisa em Português.

CONTENTS

1 INTRODUCTION

The evolutionary theory is one of the great breakthroughs in the history of science. It describes the process through which the organisms diversify over time to give rise to the incredible complexity we observe around us. The evolutionary theory provided the cornerstone that unified biology and supplied us a new point of view of the world where the whole life is deeply integrated. We all know that life is a very dynamical process at the level of individuals, with individuals being born, dying and aging constantly, but evolution shows us that this is true also at the level of the populations themselves, with species regularly appearing, getting extinguished and modifying over larger time scales.

This thesis aims to provide a theoretical contribution to the study of a specific aspect of evolution: the evolution of multicellularity and complexity. In this introductory chapter, we briefly present and discuss some relevant aspects that contextualize the work developed through the following chapters.

1.1 History of Evolutionary Theory

Until XVIII century Biology was mainly seen as static. Several factors contributed to this, namely the philosophy of Plato and Aristotle where each individual was regarded as an imperfect instance of an idealized Form. These ideal Forms were perfect and immutable and, therefore, no evolution could ever be achieved. Also the biblical view of the creation led many people to remain connected to the idea of an immutable world. This started to change soon after Carl Linnaeus introduced his taxonomic classification system in the XVIII century. Although Linnaeus himself did not propose that the species could be dynamical, the system classified the organisms according to their morphological characteristics, which led to the establishment of genetic[1] relationships between species. Besides that, the system allowed for the inclusion of species from the fossil record, which soon started to be included side by side with the living organisms. Roughly at the same time, modern Geology was emerging as a systematic field and showing that the Earth should be much older than accepted until then. With the new data in hand, especially from the fossil record, the concepts of variability in species and extinction picked up a prominent role in the prevalent ideas. In the early XIX century, Lamarck published the first true evolutionary theory. He believed that simple organisms were continuously created through spontaneous generation. These organisms would then slowly acquire a higher complexity over time driven by a natural tendency of living matter to increase in complexity. Adaptation to their environment would come from a parallel mechanism where more frequently used organs develop more and less used ones would decline. These changes would then be passed to their offspring and this continuous process would give origin to new species over long periods of time. An often cited example, provided by Lamarck himself, is the neck of a giraffe. The ancestors of the giraffe had shorter necks. Then, as a consequence of stretching them often to access high leaves in the trees, the neck would have extended over generations, culminating in the long-necked giraffe we know nowadays.

From a large set of observations and theoretical arguments, Darwin started to devise an evolutionary theory that would revolutionize our understanding of life. Darwin studied thoroughly the artificial selection made by animal breeders. He realized that different varieties arose naturally and the breeders selected them by carefully controlling the reproduction and survival rates of each variety. Aware of the work of Malthus, *An Essay on the Principle of Population* [1], he connected the idea of limited natural resources to the act of selection by the breeders. The concept of Natural Selection was born. The most adapted varieties have more successful offspring, which will dominate the future generations. For twenty years Darwin did not publish his ideas, prefering to slowly accumulate more evidence to his theory. It was only when Alfred Russel Wallace arrived independently at the same ideas that Darwin decided to publish. In 1858, Darwin and Wallace published an article together[2] containing Wallace's paper and extracts of Darwin's writing on the subject, where the authors proposed the basic ideas of natural selection. One year later,

[1] Genetic is here used in the sense of common origin, the concept of genes appeared much later. `oxforddictionaries.com` defines genetic as *relating to origin, or arising from a common origin.*

[2] *On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection* [2].

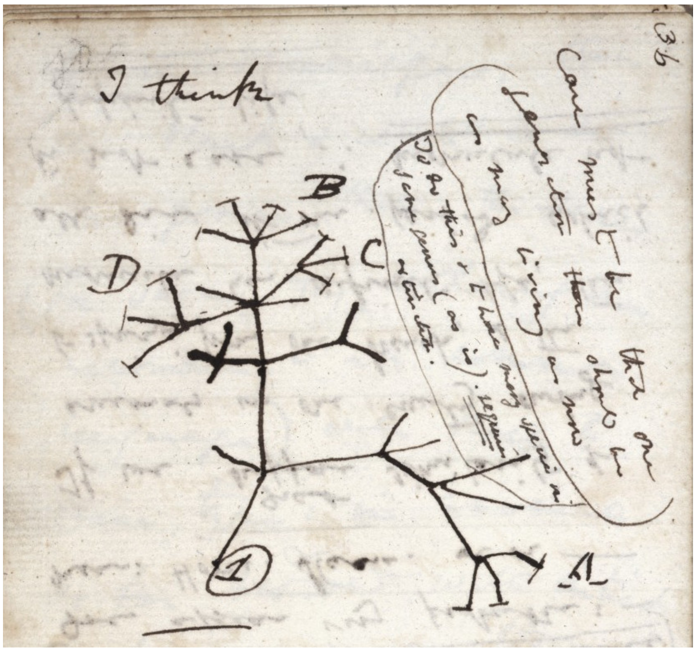
36
I think
Case must be that one generation then should be as many living as now.
To do this & to have many species in same genus (as is). requires extinction.
B
C
D
1
A

Figura 1.1: Charles Darwin's sketch of an evolutionary tree made in 1837, probably his first evolutionary tree diagram. It is in from his "First Notebook on Transmutation of Species" on display at the Museum of Natural History in New York.

Darwin finaly published his highly influential *On the Origin of Species*[3] [3]. Besides the idea of natural selection Darwin understood that there were other mechanisms in action, such as migration for example. Darwin defended the common origin of species, with the species being related to each other in a tree structure. One of such tree diagrams, probably his first, is reproduced in Fig. 1.1.

In parallel, starting in 1856, Mendel conducted a series of experiments with different varieties of pea plants which allowed him to correctly infer several of the laws of inheritance. For these studies, many consider him to be the pioneer of genetics. However, his conclusions did not have much impact at that time since most of the scientific community was unaware of his results. It was only in the beginning of XX century, with the rediscovery of his results, that real recognition was awarded to his efforts.

It was not until the beginning of XX century that another breakthrough was to happen in evolutionary theory. Starting with Ronald Fisher's paper in 1919 [4], a new integrated evolutionary theory arose from the contributions of several authors that incorporated Darwin's natural selection, Mendelian inheritance and the genetic variation of populations, besides several other mechanisms. The theory was now

[3] The full title of the book is *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.*

firmly established in biological and mathematical terms for the first time. Besides blending the theories for inheritance and selection, Modern Synthesis provided a common framework for the evolution seen in large time scales, e.g., from the fossil record, and the population dynamics and ecological aspects, observed in smaller time scales.

This theory has since been adapting and incorporated new features, specially since the development of the methods for fast genetic sequencing. Notably, it is now recognized that lateral gene transfer[4] and epigenetics[5] play a much more relevant role than previously thought although the real extent of their importance is still a subject of debate.

1.2 Evolutionary mechanisms

There are several main mechanisms responsible for evolutionary change. The primary evolutionary mechanisms are usually identified as mutation, migration, genetic drift and natural selection. For evolution to take place it is essential that the population displays variation. Selective mechanisms draw upon this variation and lead to time changes in the population.

Mutation is a source of population variation upon which selection can act. Natural selection is the mechanism Darwin is best known for. Natural selection is responsible by the adaptation of the organisms to their environment. The best adapted individuals have a higher chance of producing more offspring that achieve reproduction age, providing a higher contribution to the genotype and phenotype of future generations. Although better adapted organisms have a higher chance of shaping the future generations, biological systems are composed of a finite number of individuals and, as such, subject to stochastic fluctuations. Sometimes it is not the most adapted organism that can achieve fixation, especially when the population is small or the fitness difference is diminute. These effects are known as genetic drift and stem from the random nature of the population sampling for reproduction. Also migration plays an important role on the fate of the populations. Migration has a direct effect on the composition of the population and can contribute strongly to the success or failure of a given group.

[4] Transfer of genetic information between cells, in contrast with vertical transmission which happens from the mother cell to the daughter cells. This mechanism is relevant mainly in prokaryotic cells, which present less barriers to lateral gene transfers.

[5] Heritable phenotypical changes that do not involve changes in the genetic sequence of a cell.

1.3 Brief history of life on Earth

Earth formed together with the Solar System around 4.5×10^9 years ago [5–7]. The formation of Earth has been followed by a collision with a planetary body with the estimated size of Mars that ejected large amounts of material to the orbit of Earth [8]. This material eventually coalesced to form the Moon, which is relevant to the life on Earth namely due to the stabilizing effect it has on the orbit and orientation of our planet [9]. This stability avoids large changes of the climate over time, which would fluctuate chaotically in the absence of the Moon [9]. The early Earth is thought to have been an inhospitable place for life, during the first geological eon, the Hadean [10]. Nevertheless, recent data suggest milder conditions than initially thought, with some evidence for the existence of continental crust and oceans early in the evolution of the planet [11]. Also, a magnetic field of magnitude comparable to the recent one can be inferred from zircon data[6] from the period ranging $3.3 - 4.2 \times 10^9$ years ago [12], meaning that the Earth acquired a strong magnetic field early in its evolution.

Life is thought to have appeared in the Archaean, the following geological eon, which started 4×10^9 years ago. Exactly how life first arose is not yet known. Several concurrent theories exist that try to address this important question, but it is still largely an open topic [10, 13]. Meteoritic data has revealed that the basic elements of life were already relatively abundant in the accretion disk that gave origin to the Solar System [14–17]. Besides that, many important simple organic molecules could be formed under the conditions assumed for early Earth through abiotic processes only [16, 18].

The earliest undisputed records of life appear in the fossil record at around 3.5×10^9 years ago [19]. These fossils display a significant diversity already at play at this stage, featuring 11 different taxa, including cyanobacteria-like specimens [19]. Some data seems to imply that life may have appeared significantly sooner than that. This observation is suggested by older rocks (3.8×10^9 years old) that have isotopic ratios the carry the characteristic signatures of life [20]. More recently, even older zircon crystals (4.1×10^9 years old) containing graphite incrustments

[6] Zircon crystals are extremely resistant allowing them to survive longer than the rocks they are incrusted. As a matter of fact, some zircon crystals are older than the oldest rocks known on Earth and carry very important information from the early history of our planet in the form of small impurities trapped in the structure of the crystal. They can be dated very rigorously by measuring the proportions of uranium and lead in their structure.

that match biological isotopic signatures [21], although it is possible that some non-biological mechanism is responsible for these signatures instead.

Modern life is believed to have descended from a single ancestor, dubbed the Last Universal Common Ancestor or LUCA. The debate on how did LUCA exactly look like or in what environment did it live is still ongoing. An exciting recent study compared from a database of 6.1 million bacterial genes and identified 355 genes that are probable to have been present in LUCA [22]. These genes seem to point to our Last Universal Common Ancestor having been a microorganism inhabiting sea water thermal vents and having lived around 4×10^9 years ago. This interpretation has been challenged by other researchers [23] that point some flaws in its design which may partially invalidate the conclusions of the study. They also consider that the produced data does not support some of their claims, albeit recognizing that the study is a major step forward towards identifying the genetic mechanisms of LUCA [23].

In 1990, based on molecular data, Woese and collaborators proposed a revolutionary system of classification of life that divided life into three *domains* [24]. This system has been widely confirmed since then, superseding the previous classification that classified cellular life in Prokaryota (cell without a nucleus) and Eukaryota (cell with a nucleus) [24–26]. According to the current view of this theory, life is believed to have branched first in Bacteria and Archaea, two of the life domains proposed. The third domain, the Eukarya to which we belong, has branched from Archaea somewhere around 2×10^9 years ago. The current view of the Eukarya as branching from Archaea may lead to the inclusion of Eukarya in the latter domain, resulting in a system with only two fundamental domains of life [27]. A 2015 paper [28] describes a previously unknown group of Archaea that shares many characteristics previously associated to Eukarya only. This discovery lends credence to the theory that Eukarya originated within Archaea and provided valuable insights on the transitional forms that originated Eukarya.

The currently accepted classification systems leave out the classification of viruses and other possible noncellular forms of life, whose status remains highly debated [29,30]. The relation of these forms to the remaining life is nebulous, although some recent studies attribute them a high importance in the horizontal transport of genes among species [29, 31, 32] and eventually even in the origin of the Eukarya [29].

With life development, advanced photosynthesis has eventually appeared in a group of bacteria called cyanobacteria. These bacteria consumed water and carbon dioxide to fix carbon and obtain energy from sunlight. This process releases molecular oxygen O_2 in the environment. At first, the levels of atmospheric oxygen remained considerably low, since there were large amounts of iron in solution in the ocean that reacted with the oxygen to form insoluble iron oxides that precipitated to the bottom of the ocean. Rocks originated during that period still display these iron oxide bands. Eventually, the iron in solution in the oceans dropped below levels that could not react with all the produced oxygen anymore, and the oxygen started to stockpile in the oceans and atmosphere. As most lifeforms in this period were not adapted to oxygen, this increase in oxygen levels triggered one of the largest mass extinctions on the planet. It also provided a new generalized source of electron acceptors that allowed the development of species that perform aerobic respiration, a highly efficient form of metabolism. One lineage of free-living aerobic bacteria (identified recently as closely related to a group of modern bacteria [33]) was incorporated as an endosymbiont[7] in an early eukarya species, although it is possible that this relationship started as parasitism or predation. This symbiotic relationship evolved to a stronger relationship in which the host could not live without the endosymbiont. This endosymbiont became an organelle[8] of the Eukarya cell, the mitochondrion. A similar process has occurred in plants where the eukarya cell engulfed some form of cyanobacteria that gave origin to the chloroplasts. More recently, brown algae established such a relationship with another eukaryotic cell. These phenomena represent a huge step in evolution history since two types of cells merge to produce a new kind of cell with a more complex internal structure.

1.4 Major transitions in individuality

Major transitions in individuality are evolutionary events when the individuals transfer their own fitness to a group in order to increase the group's fitness, in such a way that the individual cannot survive or reproduce alone anymore [34–37]. This way the former individual becomes an element of the new emergent individual corresponding to the group. This new level of organization represents a major transition where the focus of selection is shifted from the parts of the group to the group itself [34–37]. This transition plays a major role in the development of complexity since it is often accompanied by a redistribution of tasks so that the parts specia-

[7] An organism living inside another for mutual benefit.

[8] Specialized intracellular structure, usually separated from the rest of the cell by a membrane.

lize in specific functions. There are plenty of examples of this kind of transitions through the history of life, the first of each probably still happening in a prebiotic context. Individual self-reproducing molecules eventually started depending on each other to reproduce giving origin to the concept of operon[9]; genes cooperate in the genome of individuals; individual cells cooperate in colonies that become individuals themselves, just to name a few. Even social insects, like ants and bees, specialize to the point that the colony can be considered an individual[10] [38,39]. A worker ant is not able to reproduce and a queen cannot subsist without the support of the colony. Neither of them can be assigned a fitness on its own.

Another type of these transitions is the permanent embedment of free bacteria in other organisms [33, 40–42], such as the mitochondria or the chloroplasts in eukaryotes. These organisms left an independent lifestyle to become an integrated part of a more complex cell. Less permanent versions of this type of integration can be found frequently across nature, in situations that stretch the concept of symbiotic relationship to its limits. Lichens are associations of fungi and algae that depend strongly on each other. Although tecnically composed of two independent species, the relationship between the fungus and the alga is so close that they are usually characterized as a species.

Of particular importance is the transition from unicellular to multicellular life. This transition is known to have independently occurred several times through the history of life, with more than 40 of these transitions identified [43, 44]. Until recently there was a dogma in evolutionary biology assuming that the multicellularity was possible only amongst eukaryotes, being the bacteria and archaea unable to achieve this stage of complexity [45]. In the early 1990's, this paradigm started to be challenged and now there exists abundant observation of multicellular behavior in bacterial colonies [45–48]. The first unequivocal fossils of multicellular organisms belong to cyanobacteria-like organisms living as early as $3.0 - 3.5 \times 10^9$ years ago [43]. Modern cyanobacteria can form multicellular aggregates with several types of differentiated cells, including cells specialized in carbon fixation, nitrogen fixation, transport and more than one type of cells specialized in group reproduction, akin to germinative line cells [49]. The cells that perform nitrogen

[9] Set of sequential genes transcribed together and subject to common regulation.

[10] As soon as 1911, William Morton Wheeler in the interesting article "The ant-colony as an organism" defended that the ant colonies and other social insects should be regarded as individuals. He argues that, among other individual characteristics, social insects display complete germ-soma differentiation [38].

fixation are unable to undergo cell division, thus being terminally differentiated and existing only to provide a service to the multicellular organism [49].

In a recent breakthrough, Ratcliff et al. could observe the transition from free cells to multicellular individuals in an experimental evolution setting [50]. They showed that simple multicellularity arises quite quickly if organisms are subjected to the right evolutionary pressures since they were able to drive a unicellular yeast to evolve to a multicellular form in a few dozen generations only. These individuals displayed several characteristics of multicellular life, as group reproduction and apoptosis (programmed cell death) [50]. The latter is a specially strong sign of multicellularity since a cell sacrifices itself for the success of the group. Also, these groups were observed to be stable even after removing the selection pressure for multicellularity from the system, without reversing to unicellular lifestyle [50]. They repeated the feat again using a unicellular alga [51], now witnessing the evolution of an alternate uni and multicellular life cycle, that leads to an appeasement of competition inside the group, due to high genetic relatedness. The team applied a selection pressure every 72 hours, by selecting the aggregates in the bottom of the culture. This gives a selective advantage to heavier cell aggregates. By around 315 generations, a new life cycle had evolved. This cycle started with a 24 hours phase where the cells actively dispersed, using their flagella to leave the aggregates. This phase was followed by a period of 48 hours during which the aggregates increased size through cell replication and almost no free cells could be found [51].

1.4.1 Role of life cycles

Life cycles shape many features of living systems such as the genetic diversity and reproduction rate. Simple binary fission life cycles are the most common process for unicellular organisms but once multicellularity comes into play there is a world of possibilities. The researchers have been puzzled by the ubiquity of certain specific life cycles within the whole universe of possible ones. The high frequency of life cycles with a unicellular stage is particularly staggering [11]. It has been long proposed that this type of life cycles helps the multicellular organism by ensuring high genetic uniformity among the constituent cells, therefore reducing genetic conflict.

In a recent work [52], Pichugin and collaborators have extensively studied the fragmentation modes of aggregates and compared the resultant growth rates. They have found that the dominant fragmentation modes are characterized either by

[11] Consider, for example, the life cycles displayed by most animals, where each organism is originated from a single egg cell which originates all the cells of the adult organism.

binary split of the progenitor into two equally sized offspring or by the production of unicellular propagules[12]. For fixed fragmentation costs, other fragmentation modes become relevant too, namely fragmentation of the aggregate in numerous small aggregates.

1.5 Evolution of cooperation

The multicellularity and the transitions in individuality are part of a broader problem which is the evolution of cooperative behavior in general. Cooperative behavior is widespread in nature. Whatever the scale we observe life, cooperation seems omnipresent. The evolution of cooperation is a hard problem to deal with in the context of Darwinian selection since it is intrinsically a competitive process. Nevertheless, there are several known mechanisms that lead to cooperation as a product of natural selection. There exists an extensive literature on this subject [53–56]. In 1963, Hamilton popularized the concept of kin selection as an explanation for the emergence of cooperation [53]. Kin selection states that as closely related organisms share many genetic traces they are more prone to help each other since their genes can be spread by the reproduction of a relative. Hamilton's rule famously states that the relatedness of the organisms should be larger than the cost imposed by cooperation for kin selection to favor cooperation. In another highly cited study, Axelrod and Hamilton used game theory and a computer tournment to show that cooperation could emerge and be stabilized as a consequence of reciprocity [54], where an organism helps another that reciprocates the deed. A review on some of these mechanisms is given by Nowak [56], which highlights five rules to the emergence of cooperation. Besides the kin selection, the review analyzes three types of reciprocity (direct, indirect and network) and the group or multilevel selection. Multilevel selection exists when selection acts at more than one level. For example, if we have groups competing among themselves while the organisms compete inside the group, say for resources. This can favor the evolution of cooperation within the group as a means to guarantee the success of the group and, consequently, of its parts.

[12] Cell whose function is to originate a new organism.

1.6 Tradeoffs

It is worth to dedicate some lines to the concept of tradeoff, which plays a central role in Biology and in this work in particular. Tradeoffs are situations when to gain in one aspect implies losing in another. They appear in any area where a decision should be made, in contexts as diverse as economy and evolution. Tradeoffs frequently arise from limited resources. Let us focus on an example to clarify this concept. If a bird lays larger eggs they will produce larger offspring with a higher survival probability. Still, producing smaller eggs allowed the bird to lay more eggs, increasing the number of offspring per breeding season. This example presents a tradeoff between size and number of eggs or, equivalently, between survivability and number of the offspring. Since the number of offspring and their survivability cannot be simultaneously optimized there should be a choice. In some cases selection may favor a compromise between these traits, while under different conditions may choose one of the extreme cases. This way tradeoffs generate variation and maintain the diversity in our planet. If it was possible to lay simultaneously large eggs (comparing to progenitor's size) and in large amounts all birds would probably do that since they could simultaneously optimize survivability and number of offspring. Thus, nature has kiwis, that lay one single giant egg, and ostrichs, that lays many small eggs (compared to body size). An extreme example that unconstrained evolution would lead to no diversity is a *Darwinian Demon*[13]. Such an organism could optimize all traits simultaneously, being able to reproduce very often and have many offspring with high survival probability as well as live long, independently of the environmental conditions. It is clear that if such an organism would be possible soon it would dominate all environments and eliminate all diversity on Earth[14].

Many types of tradeoffs exist. An often cited one is the generalist-specialist tradeoff. When an organism specializes in something, probably gets worse at performing other functions. For example, an adaptation that provides a herbivore the exact shape of mouth to eat a specific plant probably renders it less efficient eating other plants. In this case the referred herbivore can either be reasonably efficient at eating a large range of plants or very efficient eating a specific species. The

[13] Term introduced by Richard Law in 1979 to illustrate this concept [57].

[14] In a personal note, it can be argued that humans are approaching the concept of a Darwinian Demon. Although Biology does provide us many constraints, we are able to work around many of them by relying on technology. Technology allows us to colonize almost any environment. As we expand, the Earth's biodiversity is under an increased pressure.

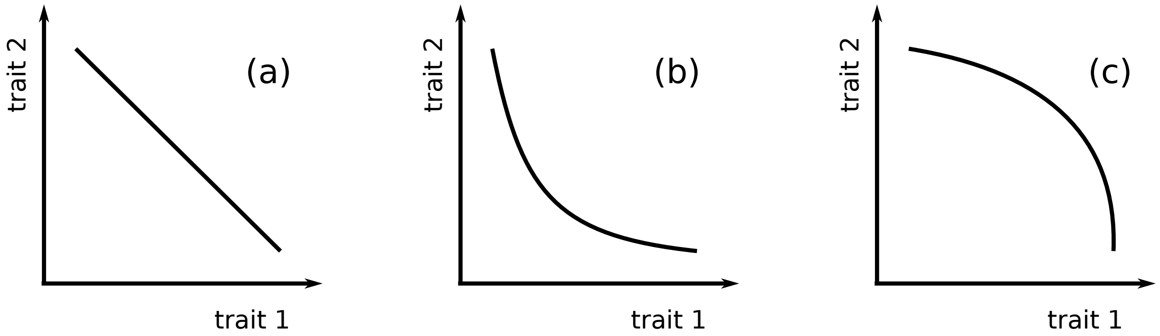

Figura 1.2: Illustration of a tradeoff between two traits, 1 and 2. a) zero curvature tradeoff; b) convex tradeoff; c) concave tradeoff.

tradeoffs exist in Biology at all scales. We referred some examples of macroscopic tradeoffs, but they exist also at the cell metabolism level. Frequently, the intermediate products of metabolism are shared among different metabolic pathways. If we increase the production of a specific metabolite, less resources are available for competing pathways. Another example is the tradeoff between rate and yield of ATP production. Fermentation of glucose, for example, is a much faster process than aerobic respiration, although aerobic respiration can achieve a much higher yield. It is possible to accelerate the process of respiration, but it implies using a larger part of the energy produced to this end implying again a smaller yield (for instance, performing active transport of some of the metabolites against the concentration gradient). Life is permanently faced with this kind of choices.

Mathematically, tradeoffs can be measured as negative correlations between traits, i.e., an increase in one trait leads to a decrease in another. The tradeoffs can display positive or negative curvature, as illustrated in Fig. 1.2. A convex tradeoff, such as the one displayed at Fig. 1.2b, tends to favor the emergence of specialization, while a concave tradeoff, like the one in Fig. 1.2c, usually favors a generalist stance. This happens because when a tradeoff is convex expressing one trait leads to a very low output of the other, while for concave tradeoffs it is possible to achieve a reasonable level of simultaneous expression in both traits.

1.7 Outline

After this introductory chapter we present and study some models, tailored to provide us an insight into several important aspects of multicellularity and complexity evolution. Specifically, chapter 2 deals with the question of the competition between different metabolic strategies in the context of multicellularity evolution.

This is an important issue since, while it is believed that multicellularity requires an efficient mode of metabolism, usually high rate and inefficient metabolism modes are selectively advantageous. A cell cannot be simultaneously highly efficient in resource usage and reproduce at a high rate, due to limiting tradeoffs. We address this problem by considering groups of cells as well as well-mixed populations. We find that efficient cells can outcompete inefficient cells for a much broader region of parameter space when grouped. This way, we help to confirm that the appearance of efficient modes of metabolisms is strongly intertwined with the advent of multicellularity. In chapter 3, we start looking into the role of the tradeoffs in the evolution of complexity. We develop a model containing generic tradeoffs and perform a statistical study on how the presence of tradeoffs leads to cell specilization in a simple multicellular organism. As a concrete example, we apply the model to the well-known system of multicellular cyanobacteria and show that it can qualitatively reproduce the observations in nature. At last, in chapter 4, we explore a mechanistic model for the dynamics of multicellular aggregates in terms of the fundamental processes that shape it. The model naturally includes the effect of tradeoffs, embedded in the fitness landscape, and allows exploring different geometries of cell aggregates. We apply the model to the study of the so-called "size-complexity rule" and obtain some surprising results. Our results reveal that the compliance of organisms with this rule depends on particular characteristics of the system, such as the geometry of the aggregate. Our work opens new ways of looking at this problem and raises a series of questions about exactly which factors determine the relation between the size and the complexity of an organism.

This ordering corresponds to the time order in which I approached these subjects during the course of my Doctorate studies and present itself as natural: an initial focus on multicellularity emergence itself, later shifted to the consequent complexity increase that follows. The different subjects approached here are further connected by the role that the tradeoffs represent in both the evolution of multicellularity and complexity.

2 COMPETING METABOLIC STRATEGIES

Highlights

A model for competition between high-rate and high-yield metabolic strategies is developed

Structured and well-mixed populations are considered

In well-mixed populations the high-yield (efficient) strain is prefered only when social conflict is absent

Structured populations favor the high-yield metabolism over a much larger range of parameters than the well-mixed populations

Analytical estimates are derived for the limits in parameter space were the efficient strain is favored, for both well-mixed and structured cases

Previous studies have suggested a link between the metabolism mode and the transition to multicellularity. In a way, efficient metabolism modes can be seen as cooperative, since the cells pay a cost in growth rate to achieve a larger total population. It is a textbook example that game theoretical models suggest that added structure favors the establishment of cooperation (for example, [58]). Pfeiffer et al. [59] studied a spatial resource-based model where similar results were found. Several observational lines of evidence seem to support this hypothesis [60, 61], including molecular evolution data [62]. In this context, our goal is to investigate the relationship between the efficiency of metabolism and the multicellularity from a point of view of a resource-based model that includes group structure since groups of cells are the basic units of multicellular life.

Heterotrophs are biological entities that process externally obtained resources to perform their activities. The cell must acquire resources which are then routed to the catabolic activities[1] responsible for the production of energy in the cell. This energy is converted into ATP, the almost universal “energy currency” of the cell, which in turn fuels whatever tasks the cell needs to fulfill. The efficiency of this process can vary greatly, from the levels of fermentation that produces around 2 ATP molecules per glucose molecule to the aerobic respiration yielding as much as 32 molecules of ATP.

In such a context, it is important to model the *rate* at which the cell acquires resources and the *yield* achieved in the process of their conversion to ATP. A high uptake rate implies that the cell can acquire large quantities of resources per time unit, while a high-yield means that the cell can produce many ATP molecules from a single glucose molecule. Ideally, an organism would aim at maximizing both yield and rate, allowing it to grow fast while making an efficient usage of the available resources. However, a set of tradeoffs exist that prevents these two traits to be simultaneously optimized. These tradeoffs can be inferred from experimental data as well as from thermodynamical arguments [59, 63–67].

In a different perspective, achieving a high-yield even at a low resource processing rate allows the total population to grow more, whereas a low-yield high-rate strategy degrades the environment fast, leading to a lower total population but allowing the strain to outcompete different strain with a lower rate. This way, a *cooperative* individual C, one that makes efficient usage of resources and thus grows slowly, is outcompeted by a *defecting* one D, that grows fast at the cost of harming the community as a whole. A C individual displays a slow metabolism, while D exhibits a fast one. The effect of these strategies on the population dynamics depicted in Fig. 2.1.

A model focusing on the efficiency and rate of resource conversion enables us to analyze the importance of the metabolism efficiency to the origin and maintenance of multicellular life. As we know, a much more efficient form of metabolism has emerged together with the appearance of free oxygen in the atmosphere, which is also coincident with some of the first known signs of multicellular life [59, 61, 68–70].

[1] Metabolism is composed of two types of activities: catabolic and anabolic ones. The catabolism is responsible for the breakdown of complex molecules to obtain energy and raw materials for the cell. Anabolism uses that energy and raw materials to synthesize complex molecules.

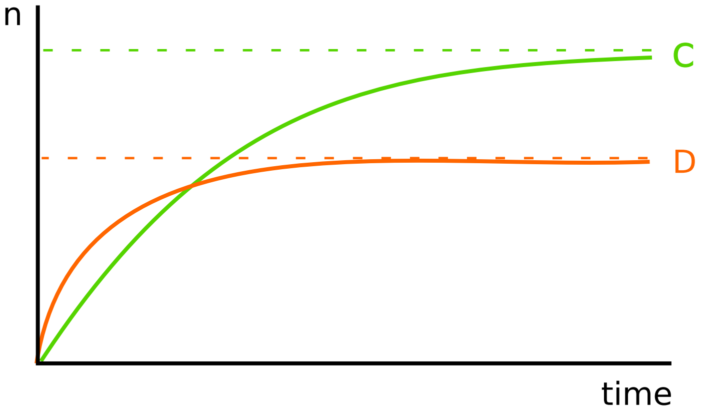

Figura 2.1: Illustration of the difference in the time evolution of populations using a high-rate low-yield strategy (C) and a low-rate high-yield strategy (D). The figure shows the population size as a function of time for a population adopting a high-rate low-yield strategy (C) and a low-rate high-yield strategy (D). As can be seen in the figure, the D strategy allows a high growth rate which saturates at a small equilibrium population, while the C strategy displays a lower growth rate but achieves a larger equilibrium population.

2.1 Model

We introduce a simple model that allows us to explore different individual behaviors by addressing two fundamental aspects of cell activity: resource uptake and processing. Essentially we model the rate at which the cell produces energy and the yield achieved. We denote the cells that achieve high-yield and low-rate by C and the cells that adopt the strategy of high-rate and low-yield by D. In the usual language of evolutionary game theory, we can describe the C cells as *cooperators* since they sacrifice part of their growing rate for the benefit of the group, and the D cells as *defectors* that take advantage of the group in order to reach a high growing rate.

The system is modelled as a discrete time process in four phases: the resource uptake, the resource processing, cell splitting and cell death. During the resource uptake phase, the cells compete for the available resources S, while during the processing phase they metabolize the captured resource. During the latter period, the cell accumulates an internal stock of energy E_j, where j is the cell index, in the form of the necessary proteins and other components required for the cell development. When the cell's internal stock of energy E_j surpasses a certain threshold E_{max}, the cell divides into two daughter cells, each inheriting half of the parent's stock. Besides these growing processes, any cell can die randomly with a uniform probability ν per time step. The details of the described cell life cycle are graphically depicted in Fig. 2.2.

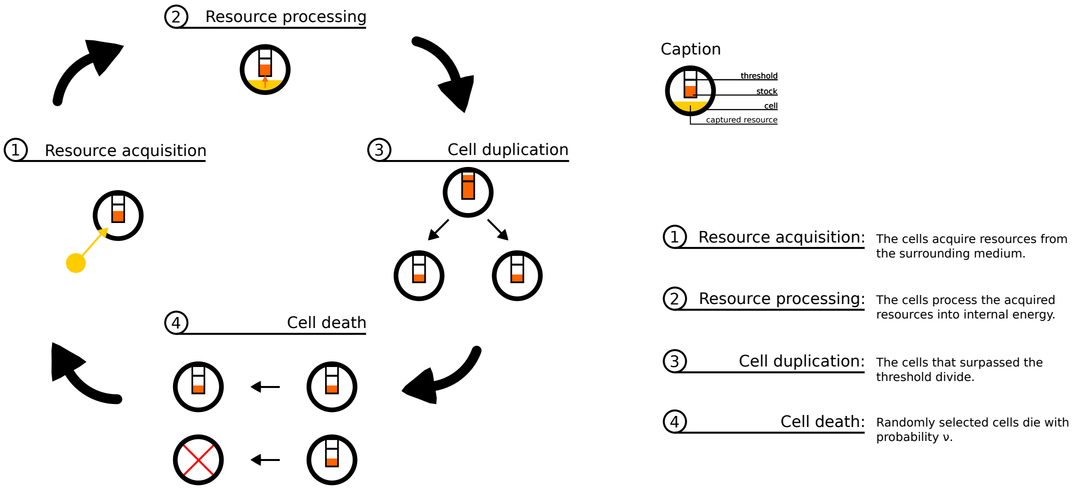
2 Resource processing
1 Resource acquisition
3 Cell duplication
4 Cell death
Caption
threshold
stock
cell
captured resource
1 Resource acquisition: The cells acquire resources from the surrounding medium.
2 Resource processing: The cells process the acquired resources into internal energy.
3 Cell duplication: The cells that surpassed the threshold divide.
4 Cell death: Randomly selected cells die with probability ν.

Figura 2.2: Illustration of the cell life cycle.

The first two phases correspond to the resource acquisition and processing and deserve further attention. The assumption of two stages is rooted in mechanistic models of the cell [71] where the resource is first transported to the interior of the cell by a set of proteins, then metabolized into a generic form of energy. The rates of resource uptake and processing are distinct [71] and the resource metabolism is under the control of multiple regulatory levels [72–74]. Nutrients are used as substrates for growth but the nutritional state also supplies signals for the cell [75]. For example, cAMP (cyclic adenosine monophosphate), synthetized from ATP, plays a major role in the regulation of the response of E. coli [76], as well as in eukaryotes [77], to different nutritional states. The design of the cell is an evolutionary choice that tunes it to act effectively within its biological niche. Our model assumes only a high-level effective description of these processes. We abstract out all the details at the intracellular level, focusing only on the traits that are relevant for our analysis of competition between metabolic regimes.

The model can then be further refined in two variations. In a basic version of the model we consider a well-mixed population, where all cells have access to the full resource supply in the system. We refer to this model as well-mixed, or occasionally homogeneous, population model. We can also add a layer of structure and organize the cells in groups, version of the model that we refer to as structured population. With structure, the cells are organized into N_G groups with population P^i $(i = 1, ..., N_G)$ individuals which split when they surpass a certain population threshold P_{max}. Upon group split, the original cells are uniformly distributed among

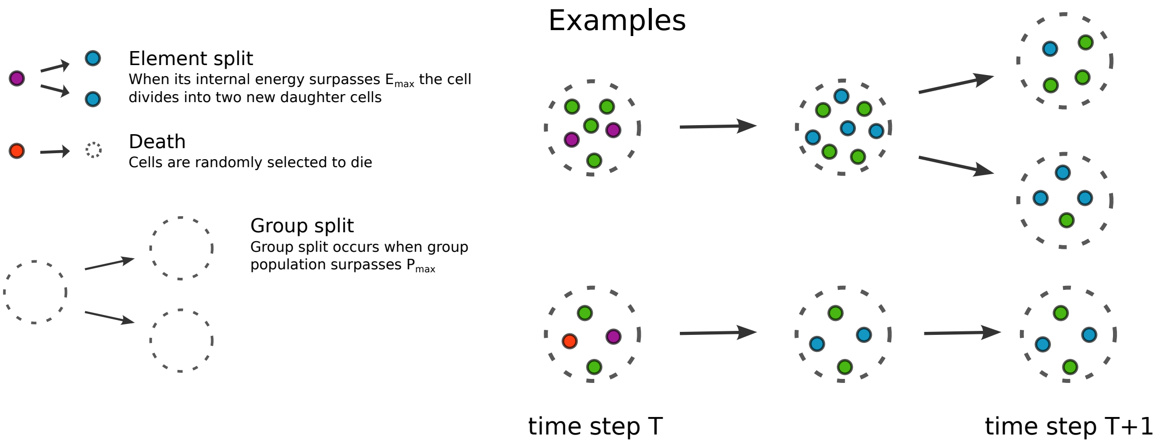
Examples
Element split
When its internal energy surpasses E_{max} the cell divides into two new daughter cells
Death
Cells are randomly selected to die
Group split
Group split occurs when group population surpasses P_{max}
time step T
time step T+1

Figura 2.3: A graphical representation of the processes of group splitting, cell division and death. Figure adapted from [79].

the daughter groups and N_G is incremented by one. The groups compete for resource, which is evenly distributed among the groups so that each of them obtains an amount of $S_G = S/N_G$ resource[2]. Notice that as N_G is a dynamical quantity, the amount of resource that each group receives depends on the specific number of groups existing at any given moment. The cells compete for the resource of the group according to the rules previously introduced. We now denote the energy stock of cell j of group i by E_{ij} ($i = 1, \dots, N_G$, $j = 1, \dots, P_i$). When cells reproduce they remain in their original group. As cells can die, a group can reach zero population. When this happens, the corresponding group is removed from the population. The rules governing the population dynamics of the structured model are illustrated in Fig. 2.3. Notice that, in both versions of the model, the total influx of resource S is a critical parameter of the dynamics, having a significant influence on the number of groups and as a result in the role of the stochasticity in the system.

As one can realize from the above description, the structured model consists of a multilevel selection model. The cells compete within the group, while the groups compete at the population level. It is in the interest of the cell to reproduce quickly and dominate the group but in the interest of the group that the average reproduction rate of its cells is high, so that the group reproduces fast. These two interests are frequently at odds with one another, therefore the two levels are conflicting. In the well-mixed model, such conflicting interests are absent and the evolution of the system is driven by the individual interests of each cell only.

[2] We could have opted by a more complex resource distribution but we chose to work with the simplest one. For a work that implements a more complex resource distribution check [78].

It remains to detail the mechanisms of resource uptake and conversion. We consider that the resource seized by each cell of type T is given by

$$S^i_T(S_G) = \frac{A_T}{A_C P^i_C + A_D P^i_D}\, S_G, \tag{2.1}$$

where T is either C or D, P^i_C and P^i_D are the number of elements of the group of type C and D, respectively, and A_C and A_D characterize the resource uptake of each cell type. The constants A_T are uptake rates and, as such, define the resource per time unit a cell can acquire. Notice that these definitions imply that $P^i_C + P^i_D = P^i$ and $S_G = P^i_C S^i_C(S_G) + P^i_D S^i_D(S_G)$, thus being consistent. The well-mixed population is treated as one global group, therefore, in this case, $S_G \equiv S$.

The resource captured by the cell is then used to increment its internal stock of energy E_{ij} by ΔE_{ij}. Considering the cell belongs to type T we have

$$\Delta E_{ij} = J_T(S^i_T). \tag{2.2}$$

The functions $J_T(S^i_T)$ describe the efficiency of conversion of resource into energy usable by the cell. The J_T functions should saturate for large values of the input resource S^i_T, with a dynamics similar to a Michaelis-Menten kinetics [59, 71]. This way we parametrize these functions as

$$J_T(S^i_T) = K_T\left(1 - \exp(-\alpha_T S^i_T)\right). \tag{2.3}$$

As one can see, K_T stands for the maximum rate of resource conversion achievable and α_T characterizes the efficiency of such process. According to our definitions, the efficient strain C should process resource in an efficient way at the expense of a lower consumption rate. On the other hand, D has a high uptake rate and metabolizes the resource quickly but it is limited to a low-yield. It is clear then that the parameters cannot take completely arbitrary values and should face some restrictions that enforce our definition. First, A_D must be larger than A_C ($A_D > A_C$), in view of the fact that the cells of strain D should be able to capture a higher fraction of the available resource. Then, since the efficiency of D should be lower than C, we have that $\alpha_D < \alpha_C$. Finally, we consider $K_D > K_C$. This choice allows cells of strain D to achieve a maximum growing rate higher than cells of strain C, which is typically the case. This possibility is particularly important since it is known that many cells can use a respiro-fermentive metabolism, concomitantly using the two alternative pathways of ATP production. Such respiro-fermentative metabolism is a typical mode of ATP production in unicellular eukaryotes like yeast [80, 81]. In such a situation, the yield of the whole process is reduced as the cell drives more resource to be metabolised inefficiently, a habitual behavior under the condition of plentiful resource. Of course, the situation $K_D < K_C$ can also be considered, but the opposite $K_D > K_C$ represents the worst scenario for the efficient strain. If it can thrive in such case, it will naturally be favoured under less harsh conditions.

We introduce shorthand notation for some dimensionless ratios that will reveal key quantities to the analysis of the system

$$\epsilon \equiv A_D/A_C, \tag{2.4}$$

$$\Delta \equiv \alpha_D/\alpha_C, \tag{2.5}$$

$$\Gamma \equiv K_D/K_C. \tag{2.6}$$

Tables 2.1 and 2.2 summarize the parameters and main quantities of the model, for the convenience of the reader.

A model in this line has been introduced by Pfeiffer et al. in 2001 [59]. In that work, the authors considered a Michaelis-Menten style function for resource acquisition and a linear function for resource processing. Adjusting the parameters, it is possible to represent different metabolic strategies, namely high-yield low-rate and low-yield high-rate strategies. They proceed to explore a spatially structured and time-continuous model, finding that the high-yield low-rate strategy dominates for low resource influx and low cell diffusion. In our work, we consider a time discrete model without spatial structure, but with the possibility of group structure. By focusing on group structure, we aim a more proper representation of the multicellularity. We also rely on different functions for resource acquisition and processing, namely our resource processing functions present a saturating behavior not present in the functions used in Ref. [59].

2.1.1 Social dilemma

We are interested in the region where the social dilemma holds, i.e., the parameter region where, in a pairwise interaction, the strain D outcompetes the strain C. Outside of this region, we can expect the strain C to be trivially favored since they have higher efficiency and are favored in a pairwise interaction. On the other hand, in the region of social dilemma D is favored in a pairwise interaction and, as such, further mechanisms are necessary if C is to outcompete D.

Parameter	Interpretation	Units
ν	death rate	time-step^{-1}
S	total available resource	resource
A_T	rate of resource capture of strain T	resource/time-step
K_T	maximum energy obtainable by strain T in a time step	time-step^{-1}
α_T	efficiency of strain T in resource processing	resource^{-1}
E_{max}	energy threshold for cell split	energy
P_{max}	threshold size for group split	dimensionless
m	migration rate (probability of each cell to migrate to a random group per time-step)	time-step^{-1}

Tabela 2.1: Summary of the parameters of the model.

Quantity	Definition	Interpretation	Units
S_G	S/N_G	resource available per group	resource
J_T	(see eq. 2.3)	rate of resource processing of strain T	energy/time-step
ϵ	A_D/A_C	fraction of the resource acquired by D in a pairwise interaction	dimensionless
Δ	α_D/α_C	relative efficiency of D and C	dimensionless
Γ	K_D/K_C	ratio between the maximum rates of resource processing of D and C	dimensionless
r	J_D/J_C	relative advantage of D over C in a pairwise interaction	dimensionless
k_T	K_T/E_{max}	maximum growth rate of strain T	time-step^{-1}
ΔE_{ij}	J_T	internal energy increase of strain T	energy

Tabela 2.2: Summary of the main quantities relevant to the analysis of the model.

For a certain amount of resource S^*, in a pairwise interaction, the amount of resource seized by strains C and D are, respectively,

$$S_C = \frac{A_C}{A_C + A_D} S^* \tag{2.7}$$

and

$$S_D = \frac{A_D}{A_C + A_D} S^*. \tag{2.8}$$

The relevant quantity to analyze the relative advantage of D over C in a pairwise competition is the ratio between J_D and J_C, which we will dub r. Therefore

$$r = \frac{J_D}{J_C} = \Gamma \frac{1 - \exp\left(-\alpha_D S_D\right)}{1 - \exp\left(-\alpha_C S_C\right)}. \tag{2.9}$$

The dilemma holds whenever $r > 1$. This ratio is a function of the amount of resource available for the competing pair S^*. We argue that $S^* \sim S/N$, where N is the population size. As we shall see from the simulations, the system evolves to a situation where S/N is always small. In this limit, the expression for r simplifies and one can obtain

$$\begin{aligned} r &= \Gamma \frac{1 - \left(1 - \alpha_D \frac{A_D}{A_C + A_D} S^*\right)}{1 - \left(1 - \alpha_C \frac{A_C}{A_C + A_D} S^*\right)} + \mathcal{O}\left(S^{*2}\right) \\ &= \Gamma \frac{\alpha_D A_D}{\alpha_C A_C} + \mathcal{O}\left(S^{*2}\right) = \Gamma \epsilon \Delta + \mathcal{O}\left(S^{*2}\right). \end{aligned} \tag{2.10}$$

Therefore, in the limit of very small resource per cell, the social dilemma region is defined by the condition

$$\Gamma \epsilon \Delta > 1. \tag{2.11}$$

Notice that, granted that the resource per cell is small, the result is independent of the exact value of resource available.

2.1.2 Simulations

We will consider a scenario where the initial population is comprised only of cells belonging to strain D. This initial homogeneous population is randomly generated and enough time is waited for it to achieve a stationary regime. At the stationary regime a population of N_{st} cells is achieved which is determined solely by the dynamics of the system and the resource provided. Once this regime is established, a single cell is replaced by a cell of type C. This stems from the assumption that, in a pool of inefficient individuals, a rare mutation will eventually arise which is more efficient than the established population. The system is then tracked until the population of strain C is either lost or fixed[3]. As we wait enough time for one of the strains to get extinguished, coexistance is never present in our results.

The fixation probability is the fraction of independent runs at which the efficient strain gets fixed. Instead of this simple fixation probability, a relative fixation is considered, dividing the absolute fixation probability by $1/N_{st}$. This allows us to analyze the probability of fixation of the newly introduced mutant with the fixation probability of a mutant under neutral selection[4], thus being a more meaningful quantity to analyze. A relative fixation probability larger than one implies the new cell is selected for, while being smaller than one means it is counterselected.

Therefore, the analysis consists of an evolutionary invasion. Extensive computer simulations are performed, spanning much of the physically meaningful parameter space. Analytical approximations are also performed for some special cases that allow it. Although providing valuable insight, deterministic approaches alone are oftentimes insufficient to study population dynamics as they disregard the stochastic effects [59]. These effects are frequently crucial to the dynamics, especially when the population is small.

The structure of the simulation procedure is described in the flowchart in Fig. 2.4.

All the C++ codes used to perform the simulations here presented are publicly available at Dryad Repository in `https://doi.org/10.5061/dryad.q6784` [82].

2.2 Analytical results for well-mixed populations

We first present the results of an analytical approximation of the model. These analytical results provide an important contribution to evaluate when the stochastic effects manifest. No population structure will be assumed in this derivation.

[3] We say a strain achieves fixation when the whole population is composed of cells of that strain.

[4] Neutral selection occurs when the strains in competition present the same fitness. In this case, all the cells have the same fixation probability.

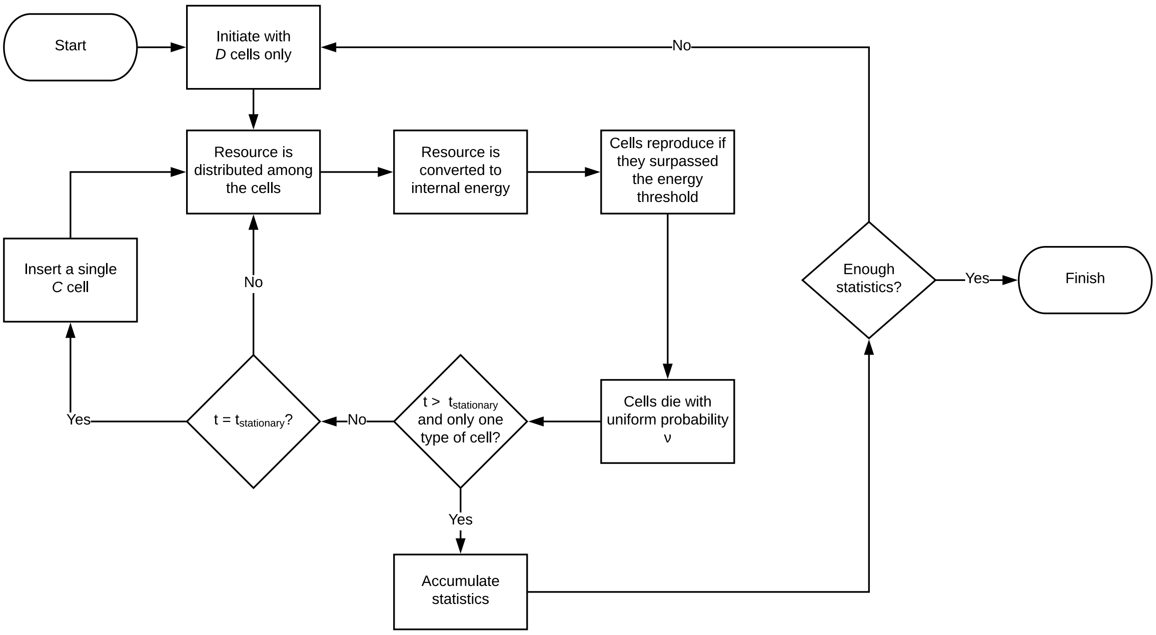
Start
Initiate with D cells only
No
Resource is distributed among the cells
Resource is converted to internal energy
Cells reproduce if they surpassed the energy threshold
Insert a single C cell
No
Enough statistics?
Yes
Finish
Yes
t = t_stationary?
No
t > t_stationary and only one type of cell?
Cells die with uniform probability ν
Yes
Accumulate statistics

Figura 2.4: Flowchart of the simulation procedure.

Let us start with a population composed by cells of strain T only. The previously described model can be captured by a time discrete equation as

$$n(t+1) = n(t) + g(n(t), S) - \nu n(t), \tag{2.12}$$

where $n(t)$ denotes the size of population at time t, $g(n(t), S)$ stands for the growth rate of strain T and the remaining parameters take the previously ascribed roles. As all cells are of type T the resource is evenly distributed. Therefore, the resource available per cell is simply $S_T = S/n(t)$. The growing rate is proportional to the energy acquired by the cell. Nevertheless, a cell only reproduces after crossing the threshold E_{max}. We can incorporate this by dividing the J_T function by E_{max}. More specifically K_T will be replaced by $k_T \equiv K_T/E_{max}$ to account for that. Inserting these aspects into the expression we find

$$n(t+1) = n(t) + k_T \left[1 - \exp\left(-\frac{\alpha_T S}{n(t)}\right)\right] n(t) - \nu n(t). \tag{2.13}$$

The system reaches equilibrium when $n(t+1) = n(t) = \hat{n}$. There are two solutions for this equation

$$\hat{n}_0 = 0 \tag{2.14}$$

and

$$\hat{n}_1 = -\frac{\alpha_T S}{\ln(1-\nu/k_T)}. \tag{2.15}$$

The second solution is only valid when $\nu < k_T$ which is intuitive as we cannot expect a nonzero equilibrium population when the death rate is always higher than the reproduction rate (please recall that k_T represents the maximum reproduction rate achievable).

We should now perform a linear stability analysis of the newfound solutions. Representing the dynamics of the model as $n(t+1) = f(n(t))$, an equilibrium solution $\hat{n}$ is stable if $|\lambda| < 1$, where $\lambda = \left.\frac{\mathrm{d}f}{\mathrm{d}n}\right|_{n=\hat{n}}$. A small perturbation on the equilibrium will be damped if $|\lambda| < 1$ but grows if the condition is not verified, leading the system out of the equilibrium.

The derivative of f reads

$$\frac{\mathrm{d}f}{\mathrm{d}n} = 1 + k_T\left[1 - \exp\left(-\frac{\alpha_T S}{n}\right) - \frac{\alpha_T S}{n}\exp\left(-\frac{\alpha_T S}{n}\right)\right] - \nu. \tag{2.16}$$

Replacing the solutions previouly found reveals us the stability regions of each solution. For $\hat{n}_1$ one has

$$\begin{aligned}\left.\frac{\mathrm{d}f}{\mathrm{d}n}\right|_{n=\hat{n}_1} &= 1 + k_T\left[1 - \exp\left(-\frac{\alpha_T S}{-\alpha_T S/\ln(1-\nu/k_T)}\right)\right.\\ &\left. - \frac{\alpha_T S}{-\alpha_T S/\ln(1-\nu/k_T)}\exp\left(-\frac{\alpha_T S}{-\alpha_T S/\ln(1-\nu/k_T)}\right)\right] - \nu\\ &= 1 + k_T\left[1 - \exp\left(\ln(1-\nu/k_T)\right) + \ln(1-\nu/k_T)\exp\left(\ln(1-\nu/k_T)\right)\right] - \nu\\ &= 1 + k_T\left[\nu/k_T + \ln(1-\nu/k_T)(1-\nu/k_T)\right] - \nu\\ &= 1 + k_T(1-\nu/k_T)\ln(1-\nu/k_T).\end{aligned} \tag{2.17}$$

Since solution $\hat{n}_1$ is only valid for $\nu < k_T$, the logarithm is always negative and $|\lambda_1|$ will remain smaller than one for the whole region of validity of the solution. Thus, the region of stability of $\hat{n}_1$ is $\nu < k_T$. Since n appears in the denominator of 2.16, we should be slightly more careful when dealing with the solution $\hat{n}_0$. For that, one can take the limit of $\frac{\mathrm{d}f}{\mathrm{d}n}$ when $n \to 0$

$$\frac{\mathrm{d}f}{\mathrm{d}n} = 1 + k_T\left[1 - \exp\left(-\frac{\alpha_T S}{n}\right) - \frac{\alpha_T S}{n}\exp\left(-\frac{\alpha_T S}{n}\right)\right] - \nu. \tag{2.18}$$

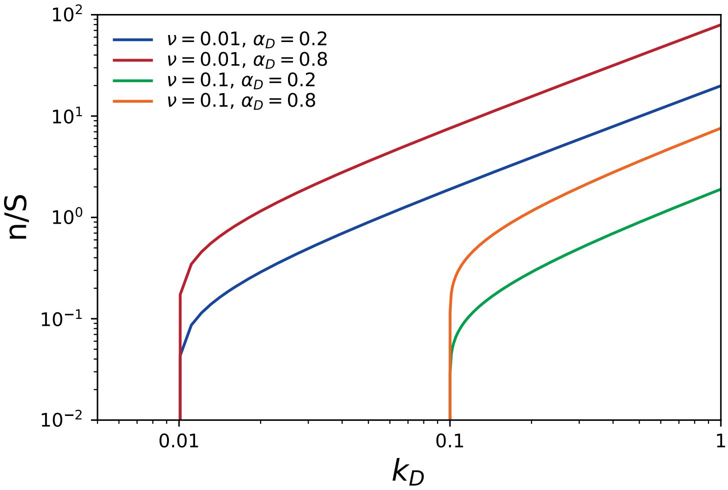

Figura 2.5: Population size as a function of k_T, for $\nu = 0.01$ and 0.1 and $\alpha_T = 0.2$ and 0.8. The scales are logarithmic.

For the stability of the solution $\hat{n}_0$ one finds

$$\begin{aligned}\lim_{n\to 0}\frac{\mathrm{d}f}{\mathrm{d}n} &= \lim_{n\to 0}\left\{1 + k_T\left[1 - \exp\left(-\frac{\alpha_T S}{n}\right) - \frac{\alpha_T S}{n}\exp\left(-\frac{\alpha_T S}{n}\right)\right] - \nu\right\} \\ &= 1 + k_T - \nu. \end{aligned} \tag{2.19}$$

From this expression one finds that the stability region of $\hat{n}_0$ is $k_T < \nu$, exactly complementary to the stability region of $\hat{n}_1$. Fig. 2.5 shows the population as a function of k_T. Notice that the population drops to zero when k_T reaches the value of ν. The graph also highlights that a higher efficiency α_T produces a higher equilibrium population. This model is a continuous approximation of the system and thus the population can be nonzero but quite small. In a discrete population model, a very small population would get extinguished quickly due to the stochastic effects that cause population to fluctuate and therefore the population would drop to zero sooner.

2.2.1 Evolutionary invasion analysis

After determining the equilibrium for cells of a single strain T we can analyze what happens if we introduce a small amount of cells of the opposite strain. For this analysis one should generalize equation 2.13 to include both strains. We obtain a system of coupled equations

$$\begin{aligned} n_D(t+1) &= n_D(t)\left\{1+k_D\left[1-\exp\left(-\alpha_D\frac{A_DS}{A_Cn_C(t)+A_Dn_D(t)}\right)\right]-\nu\right\}, \\ n_C(t+1) &= n_C(t)\left\{1+k_C\left[1-\exp\left(-\alpha_C\frac{A_CS}{A_Cn_C(t)+A_Dn_D(t)}\right)\right]-\nu\right\}. \end{aligned} \tag{2.20}$$

As both strains are present we had to reintroduce the full expression for S_D and S_C, which does not simplify anymore to S/n.

Let us consider one of the equilibria of the system 2.20. We want to know if cells of the strain C can invade the system when we introduce a small amount of cells C in an established D population. To answer this question, we should look at the stability of the solution that considers $\hat{n}_D = -\frac{\alpha_D S}{\ln(1-\nu/k_D)}$ and $\hat{n}_C = 0$, which corresponds to a population of cells of type D in equilibrium in the absence of C cells. The stability is now more complicated to calculate since we are dealing with a Jacobian matrix of the system instead of a single derivative. One must calculate the eigenvalues of the Jacobian matrix, if the absolute values of the eigenvalues are smaller than 1 the solution is stable, otherwise it is unstable. Writing the system in the form

$$\begin{aligned} n_D(t+1) &= f_D(n_D, n_C), \\ n_C(t+1) &= f_C(n_D, n_C), \end{aligned} \tag{2.21}$$

the Jacobian matrix becomes

$$J = \begin{pmatrix} \frac{\partial f_D}{\partial n_D} & \frac{\partial f_D}{\partial n_C} \\ \frac{\partial f_C}{\partial n_D} & \frac{\partial f_C}{\partial n_C} \end{pmatrix} \tag{2.22}$$

which should be evaluated at the equilibrium we want to study. The general form of entries of the Jacobian matrix for this system are

$$\begin{aligned}\frac{\partial f_D}{\partial n_D} &= 1 + k_D\left[1 - \exp\left(-\alpha_D \frac{A_D S}{A_C n_C + A_D n_D}\right)\right] \\ &\quad - \nu - n_D \frac{\alpha_D k_D A_D^2 S}{[A_C n_C + A_D n_D]^2} \exp\left(-\alpha_D \frac{A_D S}{A_C n_C + A_D n_D}\right), \end{aligned} \qquad (2.23)$$

$$\frac{\partial f_D}{\partial n_C} = -n_D \alpha_D \frac{A_D A_C k_C S}{[A_C n_C + A_D n_D]^2} \exp\left(-\alpha_D \frac{A_D S}{A_C n_C + A_D n_D}\right), \qquad (2.24)$$

$$\frac{\partial f_C}{\partial n_D} = -n_C \alpha_C \frac{A_C A_D k_D S}{[A_C n_C + A_D n_D]^2} \exp\left(-\alpha_C \frac{A_C S}{A_C n_C + A_D n_D}\right) \qquad (2.25)$$

$$\begin{aligned}\frac{\partial f_C}{\partial n_C} &= 1 + k_C\left[1 - \exp\left(-\alpha_C \frac{A_C S}{A_C n_C + A_D n_D}\right)\right] \\ &\quad - \nu - n_C \frac{\alpha_C k_C A_C^2 S}{[A_C n_C + A_D n_D]^2} \exp\left(-\alpha_C \frac{A_C S}{A_C n_C + A_D n_D}\right). \end{aligned} \qquad (2.26)$$

This greatly simplifies when applied to the equilibrium we are interested to study. After some calculations it is possible to obtain

$$J = \begin{pmatrix} 1 + k_D(1-\nu/k_D)\ln(1-\nu/k_D) & \frac{k_D}{\epsilon}(1-\nu/k_D)\ln(1-\nu/k_D) \\ 0 & 1 - \nu + k_C\left[1 - (1-\nu/k_D)^{\frac{1}{\Delta\epsilon}}\right] \end{pmatrix}. \qquad (2.27)$$

As the entrance for $\frac{\partial f_C}{\partial n_D}$ is zero the eigenvalues are trivial to calculate, corresponding directly to the diagonal entries of the matrix. They are $\lambda_1 = 1 + k_D(1 - \nu/k_D)\ln(1-\nu/k_D)$ and $\lambda_2 = 1 - \nu + k_C\left[1 - (1-\nu/k_D)^{\frac{1}{\Delta\epsilon}}\right]$. Since the region of interest is $\nu < k_D$, the logarithm in λ_1 is negative and the condition is always respected for λ_1. Let us focus on λ_2 instead. We are mostly interested in the limit $\nu/k_D \ll 1$. From expression 2.15 we can find the value of S/n. Expanding it, it is easy to see that the limit of small $\nu/k_D \ll 1$ implies also a small S/n

$$S/n = -\frac{\ln(1-\nu/k_D)}{\alpha_D} = \frac{\nu/k_D}{\alpha_D} + \mathcal{O}\left[(\nu/k_D)^2\right]. \qquad (2.28)$$

In this limit, the condition for λ_2 becomes

$$\lambda_2 < 1 \quad \Rightarrow \quad \Gamma\Delta\epsilon < 1, \qquad (2.29)$$

which is exactly the reverse condition 2.11. This shows that, up to terms of order $(\nu/k_D)^2$, the region that C can invade is the region where there is no social dilemma. Therefore, without structure, the result is trivial and the strain C cannot thrive unless they are able to overcome D in a pairwise competition. As we shall see later on, this result is in excellent agreement with the numerical simulations.

It is a simple exercise to perform an equivalent calculation, now for the invasion by the strain D of an established population of C cells. In this case, recover a similar result and conclude that the region where D can invade C is the complement of the region where C is able to invade D. Therefore, we expect that the curve given by

$$\Gamma = \frac{\nu/k_C}{1-(1-\nu/k_C)^{\frac{1}{\Delta\epsilon}}} \tag{2.30}$$

defines the curve of $P_{fix} = 1$, i.e., the relative fixation probability equals 1.

2.2.2 Coexistence solution

Besides the solutions dominated by a single strain, the system 2.20 also contains coexistence solutions. We can start by dividing the equations by $\hat{n}_D$ and $\hat{n}_C$, respectively, given that we are not interested in the solutions that include extinction of one of the populations

$$\begin{aligned} 0 &= k_D\left[1-\exp\left(-\alpha_D\frac{A_D S}{A_C\hat{n}_C + A_D\hat{n}_D}\right)\right] - \nu, \\ 0 &= k_C\left[1-\exp\left(-\alpha_C\frac{A_C S}{A_C\hat{n}_C + A_D\hat{n}_D}\right)\right] - \nu. \end{aligned} \tag{2.31}$$

With some more manipulations, one can find

$$\begin{aligned} 1-\nu/k_D &= \exp\left(-\alpha_D\frac{A_D S}{A_C\hat{n}_C + A_D\hat{n}_D}\right), \\ 1-\nu/k_C &= \exp\left(-\alpha_C\frac{A_C S}{A_C\hat{n}_C + A_D\hat{n}_D}\right), \end{aligned} \tag{2.32}$$

from which it follows that

$$\begin{aligned} A_C\hat{n}_C + A_D\hat{n}_D &= -\alpha_D\frac{A_D S}{\ln(1-\nu/k_D)}, \\ A_C\hat{n}_C + A_D\hat{n}_D &= -\alpha_C\frac{A_C S}{\ln(1-\nu/k_C)}. \end{aligned} \tag{2.33}$$

Finally, we can divide both equations by A_C so that the dependence appears in the dimensionless parameter ϵ

$$\begin{aligned}
\hat{n}_C + \epsilon \hat{n}_D &= -\alpha_D \frac{\epsilon S}{\ln(1-\nu/k_D)}, \\
\hat{n}_C + \epsilon \hat{n}_D &= -\alpha_C \frac{S}{\ln(1-\nu/k_C)}.
\end{aligned} \tag{2.34}$$

This system implies not a single solution but rather a family of solutions, shown in Fig. 2.6. If the population has only a single strain we recover expression 2.15. Matching both equations we find that the coexistence solution requires an extra constraint in the parameters

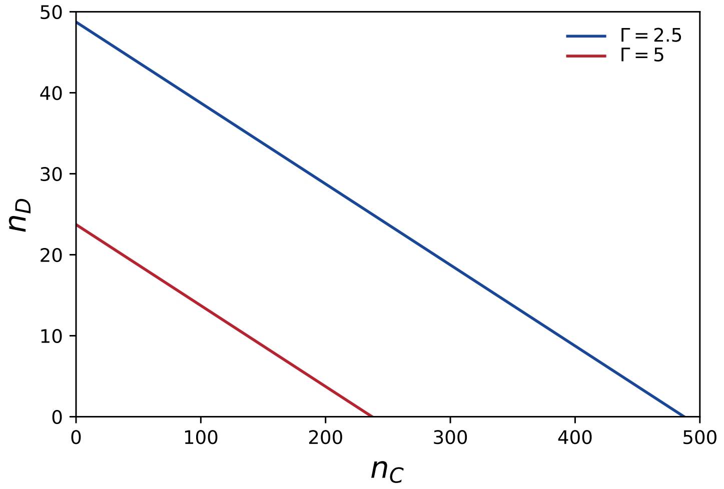

Figura 2.6: n_C and n_D population sizes for coexistence solution. All points along the lines represent coexistence solutions. The parameters are $k_D = 0.5$, $\nu = 0.01$, $S = 25$, $\alpha_C = 1$, $\epsilon = 10$ and $\Gamma = 2.5$ and 5.

$$-\alpha_C \frac{S}{\ln(1-\nu/k_C)} = -\alpha_D \frac{\epsilon S}{\ln(1-\nu/k_D)} \tag{2.35}$$

that leads to

$$\Delta\epsilon = \frac{\ln(1-\nu/k_D)}{\ln(1-\nu/k_C)}. \tag{2.36}$$

In the small ν/k_T regime, we recover once more the expression $\Gamma\Delta\epsilon = 1$. Intuitively, it shows that the equilibrium exists exactly over the line that separates the region dominated by C from the region dominated by D.

We can analize the stability of the solution by calculating the eigenvalues of the Jacobian. We can eliminate the dependence in A_C and A_D in favor of ϵ from the general Jacobian described by Eqs. 2.23-2.26

$$\frac{\partial f_D}{\partial n_D} = 1 + k_D\left[1 - \exp\left(-\alpha_D \frac{\epsilon S}{n_C + \epsilon n_D}\right)\right] \\ - \nu - n_D \frac{\alpha_D k_D \epsilon^2 S}{[n_C + \epsilon n_D]^2} \exp\left(-\alpha_D \frac{\epsilon S}{n_C + \epsilon n_D}\right), \tag{2.37}$$

$$\frac{\partial f_D}{\partial n_C} = -n_D \alpha_D \frac{\epsilon k_C S}{[n_C + \epsilon n_D]^2} \exp\left(-\alpha_D \frac{\epsilon S}{n_C + \epsilon n_D}\right), \tag{2.38}$$

$$\frac{\partial f_C}{\partial n_D} = -n_C \alpha_C \frac{\epsilon k_D S}{[n_C + \epsilon n_D]^2} \exp\left(-\alpha_C \frac{S}{n_C + \epsilon n_D}\right) \tag{2.39}$$

$$\frac{\partial f_C}{\partial n_C} = 1 + k_C\left[1 - \exp\left(-\alpha_C \frac{S}{n_C + \epsilon n_D}\right)\right] \\ - \nu - n_C \frac{\alpha_C k_C S}{[n_C + \epsilon n_D]^2} \exp\left(-\alpha_C \frac{S}{n_C + \epsilon n_D}\right). \tag{2.40}$$

Replacing $n_C + \epsilon n_D$ and performing some algebra, this can be further simplified to

$$J = \begin{pmatrix} 1 - k_D L_D & -\frac{1}{\epsilon} k_C L_D \\ -\epsilon k_D L_C & 1 - k_C L_C \end{pmatrix} \tag{2.41}$$

where $L_T \equiv \frac{n_T}{S\alpha_T}\left[\ln\left(1 - \nu/k_T\right)\right]^2 \left(1 - \nu/k_T\right)$. The eigenvalues of this matrix are $\lambda_1 = 1$ and $\lambda_2 = 1 - (k_C L_C + k_D L_D)$. The first eigenvalue is expected, since the solutions are free to move along the curve defined by equation 2.34. The second eigenvalue tells us that the solution exists whenever

$$-1 < \lambda_2 < 1 \Rightarrow -1 < 1 - (k_C L_C + k_D L_D) \wedge 1 - (k_C L_C + k_D L_D) < 1 \tag{2.42}$$

$$\Rightarrow k_C L_C + k_D L_D < 2 \quad \wedge \quad k_C L_C + k_D L_D > 0 \tag{2.43}$$

The second condition is always true, so it remains to determine $k_C L_C + k_D L_D < 2$. In the regime of small ν/k_T, L_T becomes

$$L_T = \frac{n_T}{S\alpha_T}\left(\nu/k_T\right)^2 + \mathcal{O}\left[(\nu/k_T)^3\right] \tag{2.44}$$

so

$$k_C \frac{n_C}{S\alpha_C} (\nu/k_C)^2 + k_D \frac{n_D}{S\alpha_D} (\nu/k_D)^2 < 2. \tag{2.45}$$

As we saw previously, in the worst case scenario, $S/n_T \approx \nu/(k_T \alpha_T)$, therefore we can find $\nu < 1$, which is always satisfied since most of the interesting regimes consist of $\nu \ll 1$.

Nevertheless, we cannot expect to find this solution in the simulations. Oscillations along the line are not supressed and, therefore, the population will eventually reach the points $n_C = 0$ or $n_D = 0$, where one of the strains extinguishes. Also, the coexistence is found only for a line in parameter space which makes this solution too restrictive to be of practical relevance.

2.3 Analytical results for structured populations

Some estimates can be given regarding the invasion of an established population of one strain by the other, even in the case of structured populations. We will look at the groups as emergent individuals and consider only pure groups of cooperators and defectors. This way, we can provide an estimate for the group growth rate as

$$J_T^G = P_i\, k_T \left[1 - \exp\left(-\alpha_T \frac{S}{N_G P_i}\right)\right] / P_{max} \tag{2.46}$$

where J_T^G stands for the reproduction rate of a group constituted by strain T cells, N_G for the number of groups and P_i for the population of the group. As the groups are considered to be uniform, the resource available per cell is simply $S/(N_G P_i) \sim S/P_{total}$. Furthermore, the P_{max} normalization stems from the fact that a group only splits when its size reaches P_{max} cells. With this in mind, one could introduce a quantity r_G, defined as the ratio of reproduction rates of strain D and C groups

$$r_G \equiv \frac{J_D^G}{J_C^G} = \frac{\frac{P_i\, k_D}{P_{max}} \left[1 - \exp\left(-\alpha_D \frac{S}{N_G P_i}\right)\right]}{\frac{P_i\, k_C}{P_{max}} \left[1 - \exp\left(-\alpha_C \frac{S}{N_G P_i}\right)\right]} = \Gamma\, \frac{1 - \exp\left(-\alpha_D \frac{S}{N_G P_i}\right)}{1 - \exp\left(-\alpha_C \frac{S}{N_G P_i}\right)}. \tag{2.47}$$

In the derivation of the previous equation, we compared groups of the same size. As formerly discussed, the small resource limit is a relevant limit. In this limit the expression reduces to

$$r_G = \Gamma \frac{1 - \left(1 - \alpha_D \frac{S}{N_{total}}\right)}{1 - \left(1 - \alpha_C \frac{S}{N_{total}}\right)} + \mathcal{O}\left[(S/N_{total})^2\right] = \Gamma\Delta + \mathcal{O}\left[(S/N_{total})^2\right]. \tag{2.48}$$

The curve defined by $r_G = 1$ reveals the limiting value of the region where C groups have advantage over D groups. This is a very interesting result as it shows us that, by grouping together, the C cells get rid of a factor ϵ in their disadvantage over D and can thrive over a much wider range of parameters.

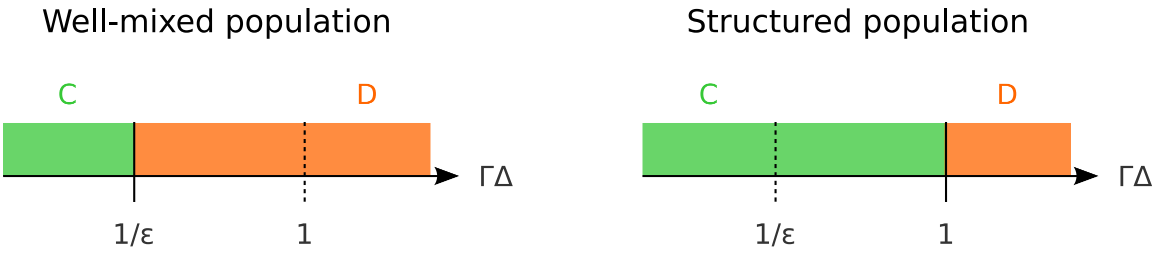

Figura 2.7: Representation of the parameter regions expected to be dominated by strategies C and D, for the cases of well-mixed and structured populations.

Based on the results obtained until now, our analytical estimates of the process allow us to expect three regions with qualitatively different behaviors. When $\Gamma\Delta\epsilon < 1$ there is no social conflict and strain C is expected to be prefered by natural selection in any situation. For the intermediate case, where $1/\epsilon < \Gamma\Delta < 1$, the strain C is predicted to be disadvantageous in well-mixed population but superior while competing in the context of structured populations. Finally, a region defined by $\Gamma\Delta > 1$ where D is always selected for, regardless of the existence of structure. Fig. 2.7 shows a graphical representation of these regions. Interestingly, the intermediate region is highly relevant since the rate-yield tradeoff guarantees that increasing Γ should lead to a decrease in Δ, keeping the product relatively unchanged. One strain can choose to have a high-yield or high-rate strategy, but not both simultaneously.

2.4 Simulation results

Having presented some analytical calculations, we now turn to the simulations of the system. Fig. 2.8 depicts the relative fixation probability of a single C cell in a

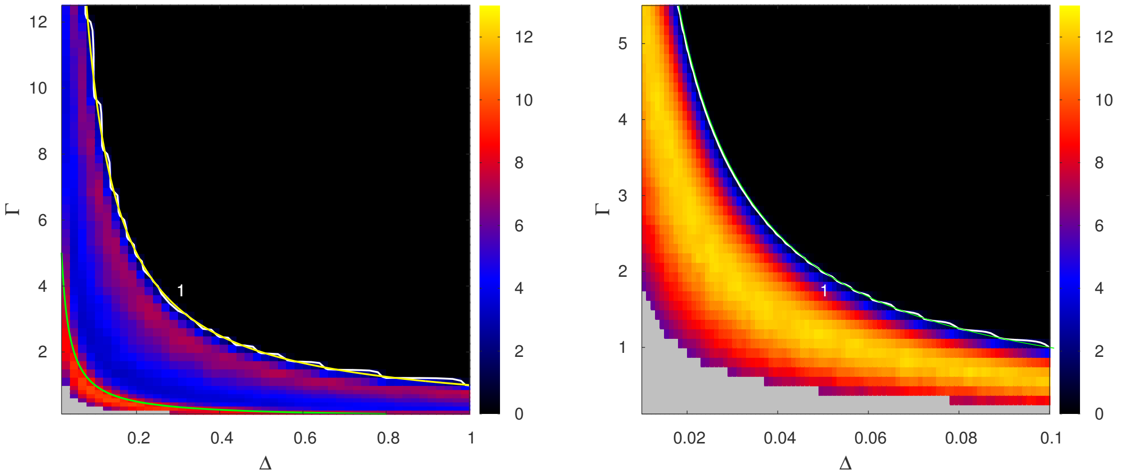

Figura 2.8: Relative fixation probability. In the plot the relative fixation probability is shown in terms of $\Gamma = K_D/K_C$ and $\Delta = \alpha_D/\alpha_C$. Left panel: structured populations, right panel: homogeneous populations. The white thick lines denote the isocline where the relative fixation probability of a single cooperator is equal to one, meaning that its fixation probability is the same of a random individual under neutral selection. Above the isocline the cooperative strategy is counter-selected (dark region), whereas under the line it is selected for. Yet, the green line corresponds to the line delimiting the social dilemma regime, obtained by making $r = 1$ in Eq. 2.11. This line overlaps the curve defined by Eq. (2.30). The yellow line in the structured population denotes the line defined by Eq. (2.48) for $r_G = 1$. The grey region denotes extinction of D population before the C cell is introduced. The other parameter values are resource amount $S = 25$, death rate $\nu = 0.01$, group carrying capacity $P_{\max} = 10$, internal energy for splitting $E_{\max} = 10$, and $A_D = 10$. The data points correspond to 40 distinct populations and for each population $10\,000$ independent runs were performed. Figure adapted from [79].

previously established population of strain D. This is shown as a function of the ratios $\Gamma \equiv K_D/K_C$ and $\Delta \equiv \alpha_D/\alpha_C$. $\epsilon \equiv A_D/A_C$ is fixed at 10, a typical empirical value from *Saccharomyces cerevisiae* populations [81]. Let us first focus on the well-mixed populations. These populations display two well-defined regions, separated by the curve defined by Eq. 2.11. Below this curve the relative fixation probability of strain C is larger than one, while above it is smaller. This reflects the fact that strain C is only able to outcompete D at the population level in the region where C already manage to succeed in a pairwise interaction. The isocline that marks relative fixation probability equal to one, in white, is in excellent agreement with the analytical expectation, in green. On the other hand, the structured population exhibits a much larger region with relative fixation probability larger than one. This region is interesting because, although the social conflict exists and cells of type C are disfavored in direct competition with D, C still achieves a high relative fixation probability. This happens due to the conflicting interests of the cell and group. In the large region of parameters between the green and white lines, the group effect can overcome the individual cell interests. As estimated analytically, this region is delimited by the condition $\Gamma\Delta = 1$. The high degree of consistency between this line (red curve in the graph) and the isocline that highlights the $P_{fix} = 1$ is noteworthy.

Fig. 2.9 represents two cuts of Fig. 2.8 with constant Γ. It allows a more detailed view over the behavior of the relative fixation probability. The fixation probability presents two peaks, whereas the homogeneous case has only one. The peak in homogeneous population, as well as the first peak in the structured population, can be attributed to the absence of social dilemma in the region. The second hump, present only in the structured populations, owes its existence to the group selection, which favors more efficient constituent cells whenever $\Gamma\Delta < 1$. For a demonstration of that, please see the dashed vertical lines denoting $\Delta = 1/\Gamma$. After this threshold, the relative fixation probability abruptly falls to zero and the C cells do not have a chance of success anymore, even accounting for group structure effects. Our analytical approximations provide an explanation for the threshold values of the regions each strain dominates but supply no guidance to explain the details of the dynamics in between the limiting values. To understand the dynamics here we need to consider the interplay between the dynamics of group formation with within-group dynamics. It may seem counterintuitive that, in a first stage, the relative fixation probability of strain C increases with Δ since higher Δ entails more efficient strain D cells. For explaining this result we need to take into account how

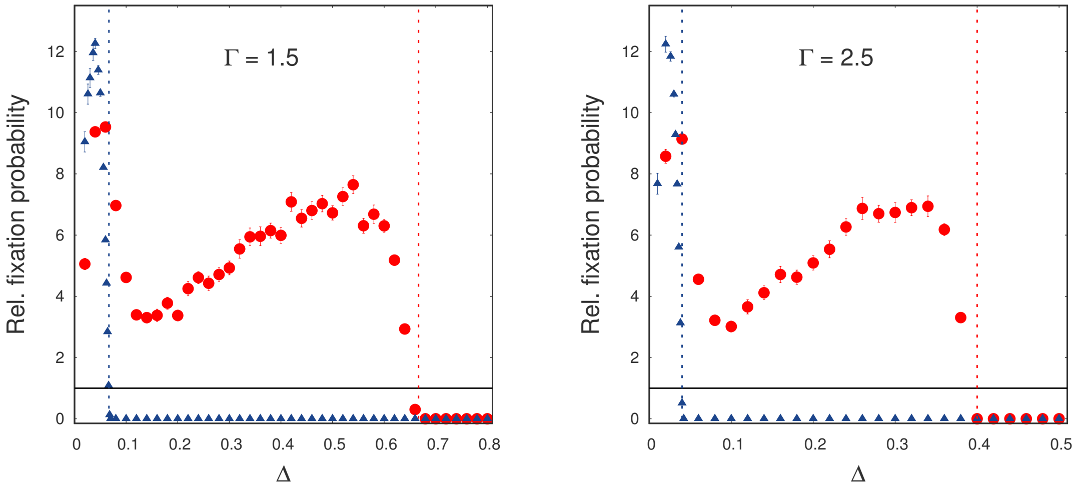

Figura 2.9: Relative fixation probability as a function of Δ for fixed values of Γ. The blue triangles show the simulation outcomes for homogeneous populations, whereas the red dots for structured population. The data points correspond to 40 distinct populations and for each population $100\,000$ independent runs were performed. The black horizontal line denotes relative fixation probability 1. The vertical dotted blue line marks the limit of the social conflict region, Eq. 2.30, whereas the vertical dotted red line designates the limit of the region for the group conflict, Eq. 2.48. The parameter values are $S = 25$, $\nu = 0.01$, $P_{\max} = 10$, $E_{\max} = 10$, $A_D = 10$, and $A_C = 1$. Figure adapted from [79].

both local competition for resource and group expansion are affected. On the one hand, although a more efficient D (larger Δ) leads to an increased strength of local competition for resource as experienced by C, on the other hand it also allows a faster expansion of the group containing the invading C cell, thus favoring group division and spreading the C trait. Thus, the net advantage of the efficient strain comes from the net outcome of these two competing mechanisms.

2.4.1 Within-group stochastic dynamics

Having analyzed the effects of the relative efficiencies and yields on the population dynamics, it is important to establish the effect of maximum group size $P_{\max}$. The $P_{\max}$ size considerably affects the stochastic effects of the within-group dynamics. The results of the simulation as a function of $P_{\max}$ and Δ are presented in Fig. 2.10. First, there is a maximum value of Δ beyond which the fixation probability is always smaller than one. As evinced by the vertical green line in the plot, this corresponds to the value of Δ beyond which a pure C group has a smaller growth rate than a pure D group. Whenever Δ is not too small there is another limit for the invasion of D by a C mutant. This is observed as a horizontal line with $P_{fix} = 1$ at $P_{max,c} \approx 17$.

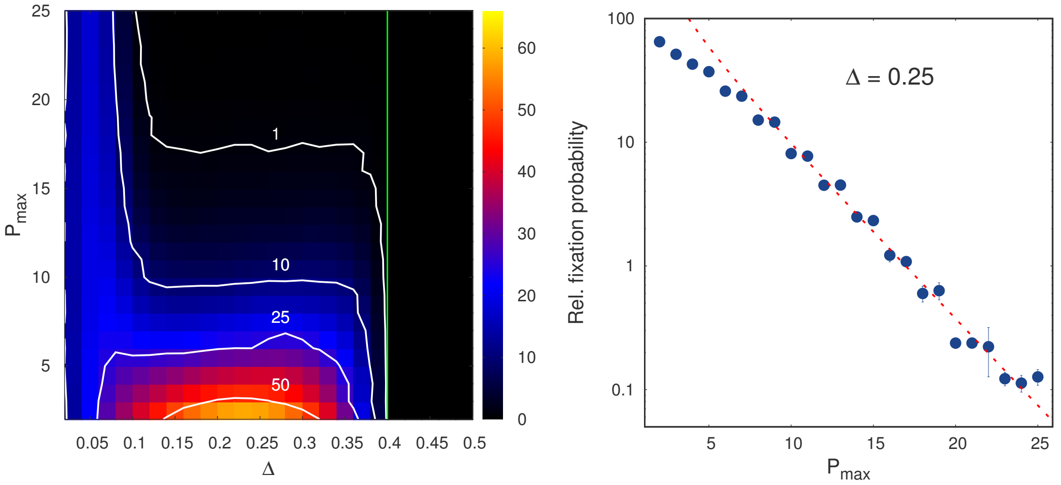

Figura 2.10: Relative fixation probability as a function of $P_{\max}$. The left panel shows the relative fixation probability in terms of the carrying capacity $P_{\max}$ and the ratio Δ. The right panel represents how the relative fixation probability changes with $P_{\max}$ in a given Δ. The thick white lines in the left panel correspond to isoclines and the green vertical line denotes the condition $\Gamma\Delta = 1$. The red dotted line in the right panel shows a fit to a Moran invasion process expression. The parameter values are $S = 25$, $\nu = 0.01$, $E_{\max} = 10$, $A_D = 5$, and $\Gamma = 2.5$. The data points correspond to 40 distinct populations and for each population $100,000$ independent runs were performed in left panel (right panel). Figure adapted from [79].

To explain these results we need to include within-group dynamics. Note that the analytical result for the regions dominated by C or D groups was obtained assuming no internal group dynamics. For groups larger than the threshold size $P_{max,c}$, the internal dynamics of the group should dominate over the intergroup dynamics, leading to the almost certain loss of C before it can control one group. To analyze this hypothesis we study in more detail a slice of the graph with constant Δ, shown in the right panel of Fig. 2.10. A coarse grained model can be established to address this. If the characteristic timescales at which the within- and inter-group dynamics take place are very different we can separate the fixation probability as the product of the fixation probability of one C cell inside a group by the fixation probability of a C-only group in the population. When the group size is small we expect the timescales associated with these dynamics to be relatively close because a few cell divisions lead to a group split. The most extreme case is the group of size two, where each cell division causes the group to split. Following this line of thought, for relatively large P_{max}, we should be able to decompose P_{fix} as

$$P_{\mathrm{fix}} \approx P_{\mathrm{fix,w}}\, P_{\mathrm{fix,p}} \tag{2.49}$$

where $P_{\mathrm{fix,w}}$ stands for the probability of fixation of one C cell in a group and $P_{\mathrm{fix,p}}$ for the probability of fixation of the group in the population. In the parameter region defined by $1/\epsilon < \Gamma\Delta < 1$ we can be certain that $P_{\mathrm{fix,w}} < 1$ and $P_{\mathrm{fix,p}} > 1$, i.e., the C strain cells are in a disadvantageous position in the within-group competition but the pure C groups enjoy higher fixation probability at the population level. From now on, we apply a series of rough approximations that should enable us to obtain a qualitative idea of what is happening. $P_{\mathrm{fix,p}}$ should not depend strongly on P_{max} since it mainly influences the stochastic dynamics inside the group, so we will take it as constant. Furthermore, we can approximate the within group dynamics by a Moran process and assign the probability of the cell achieving fixation to the corresponding expression (for a description of the Moran process, please see Appendix B). The fixation probability of a mutant in a Moran process under constant selection is given by

$$P_{\mathrm{fix,w}} = \frac{1-R}{1-R^N}, \tag{2.50}$$

where R stands for the relative fitness of the mutant and N is the population size. Therefore, we can approximate the total relative fixation probability of a C strain cell as

$$P_{\text{fix}} \approx P_{\text{fix,p}} \, \frac{1-R}{1-R^{f\,P_{\max}}}, \tag{2.51}$$

where $P_{\text{fix,p}}$ was introduced to account for the group fixation and f encodes the fraction of P_{max} that actually exists in a group. The product $f\,P_{\max}$ is consequently an effective population of the group as seen by the invading cell. Note that, in the real process, the relative fitness of the cell is dependent on the exact composition of the group. As such, this model provides us an effective view of the process and the population and relative fitnesses obtained from it should be looked upon in this perspective. As the group size P_i is usually in the range $[P_{\max}/2, P_{\max}]$[5], we expect to find f between 0.5 and 1. Fitting the above expression to the data we find $R = 0.66$, $f = 0.77$ and $P_{\text{fix,p}} = 455.2$. This fit is displayed as a red dotted line in the right panel of Fig. 2.10. As we expected, the fit does not describe well the points for small $P_{\max}$ but adjusts naturally to the end of the curve. The parameters that emerge from the fit are consistent with our intuition of the problem. We expect $P_{\text{fix,p}}$ to be quite large since for small group size the total fixation probability is large and $P_{\text{fix,w}}$ does not help, being constrained to less than 1. An effective group population of $0.77P_{\max}$ is quite natural and a fitness ratio $R = 0.66$ is also acceptable taking into account that, in this region of Δ, a C cell is always disfavored in direct competition with a D cell and the relative fitness should therefore be smaller than one. This study confirms that we can ascribe the fall in fixation probability of the invading cell to the within-group dynamics. In a larger group, the invading cell faces an increased difficulty to achieve fixation within the group. Also interesting is the possibility of testing different lifecycles. A lifecycle that includes a unicellular stage should greatly promote the fixation of the strain C in the region where C groups are selected for since this lifecycle promotes the existence of uniform groups and decreases the internal competition.

2.4.2 Resource consumption rate

The resource consumption rate is characterized by the parameter ϵ. As the cells partition the entirety of the resource in every time step, ϵ is the only relevant quantity and there is no regime where we need to know A_D or A_C individually. Figure 2.11

[5] This happens because the groups grow until they reach $P_{\max}$ size and then split in two.

unveils how the resource consumption rate affects the fate of the invading efficient strain. The relative fixation rate dependence with Γ and ϵ is probed. Once again one can see that the line of $P_{fix} = 1$ is quite well approximated by the curve $\Gamma\Delta = 1$. One can realize that the boundary between the regions dominated by C and D is almost independent of ϵ as predicted by that analytical estimation, although the details of the fixation probability shape a much more complex picture. Intuitively, lower values of ϵ entail larger probability of success for the C invader, since more resources are available to it. Although a very high D consumption rate leads to a large advantage of D inside the groups, it does not provide any advantage once uniform groups have been established. For large ϵ the probability of fixation becomes essentially constant: increasing ϵ above 5 does not lead to any substancial variation.

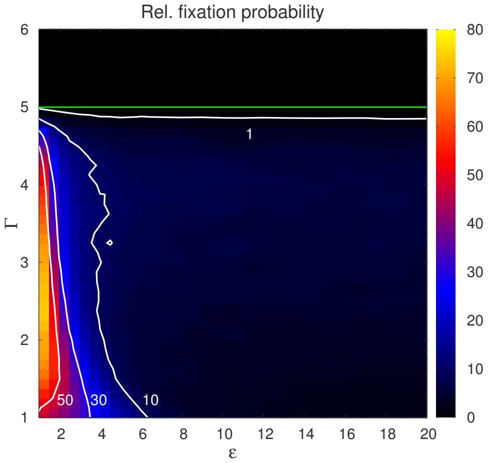

Figura 2.11: The effect of the consumption rate. Heat map of the relative fixation probability in terms of the ratios Γ and ϵ. The parameter values are $S = 25$, $\nu = 0.01$, $P_{\max} = 10$, $E_{\max} = 10$ and $\Delta = 0.2$. The thick white lines correspond to isoclines. The horizontal green line denotes the $\Gamma\Delta = 1$ curve. The data points correspond to an average over 10 distinct populations and for each population $10,000$ independent runs. Figure adapted from [79].

2.4.3 Migration between groups

Finally, for the sake of completeness, we analyze the effect of introducing the possibility of migration of cells between groups. The migration between groups should disfavor the fixation of the strain C. This happens because it becomes more difficult to achieve and maintain groups of one single cell type, keeping the internal group competition high. Now, any cell can move to a randomly chosen group with a certain probability m per time step. This implements the so-called island model of migration [83]. Allowing migration changes the effective population size and in the high migration rate limit the structured population is expected to behave similarly to its well-mixed counterpart. Fig. 2.12 depicts the effect of migration in the resultant dynamics. As one can see, there are two different behaviors. When the parameters are chosen such as the condition $\Gamma\Delta\epsilon < 1$ is valid the result remains basically unaffected by migration since the strain C is favored both within groups and between them. On the other hand, when $\Gamma\Delta\epsilon > 1$ the fixation probability remains roughly constant, until a threshold migration rate is achieved for which the strain C is not able to invade the population anymore. Beyond this value, the fixation probability is basically zero. The exact point at which this threshold occurs depends on the maximum population of a group P_{max}, with larger P_{max} leading to the colapse of C at a smaller migration rate. Beyond the transition the result is the same as obtained in the homogeneous population since for high migration rates the population becomes effectively a well-mixed population.

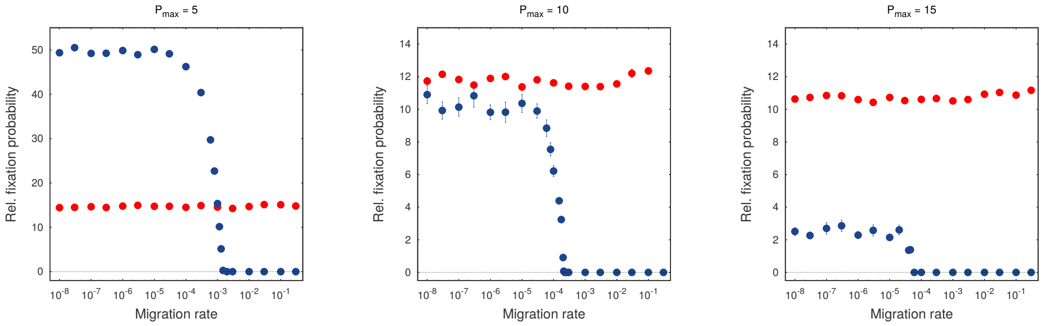

Figura 2.12: Effect of migration on the relative fixation probability. The relative fixation probability is plotted as a function of the migration rate m for three distinct values of carrying capacity $P_{\max}$. The parameter values are $S = 25$, $\nu = 0.01$, $E_{\max} = 10$, $\epsilon = 10$, $\Gamma = 1.5$ and $\Delta = 0.5$ (blue points) and $\Delta = 0.04$ (red points). Figure adapted from [79].

2.5 Conclusion

We have proposed and analyzed a model for competition of organisms with different metabolic strategies over a single limiting resource. We parametrize the fundamental aspects of the metabolism in a simple way, focusing on the rate of resource acquisition and yield achieved. The model incorporates multilevel selection in a natural way, making it an adequate tool to study the relation between the metabolism mode and the evolution of multicellularity.

Our results evince the importance of group formation for the establishment of efficient modes of metabolism. In fact, it is known that multicellular organisms possess efficient modes of metabolism much more often than unicellular ones. When groups of inefficient cells emerge inside multicellular organisms the organism is often led to disruption, as in the case of cancer. Cancerous cells possess an effect, known as Warburg effect [84], where the cells shift from their usual efficient metabolism to highly inefficient high-rate modes, incompatible with the long term sustainability of the organism.

This work launches the foundations for an approach that can be extended in many directions. One such application was studied by our group in [85], with the study of the coexistence of microorganisms with different metabolic strategies in a well-mixed environment, where one of the strains produces a toxin as a by-product of the metabolism. Another promising line of work is a formalization and extension of the arguments provided in Sec. 2.4.1 so that an effective population size could be studied as a function of the metabolic properties. Several other aspects still remain to be addressed, such as considering groups with a dynamical maximum size stemming from an underlying evolutionary mechanism, or different lifecycles or group structures.

The main results presented in this chapter have been published in the article:

- Ref. [79]: Competing metabolic strategies in a multilevel selection model, André Amado, Lenin Fernández, Weinei Huang, Fernando F. Ferreira, and Paulo R. A. Campos. Royal Society Open Science (2016), DOI: 10.1098/rsos.160544

The code used to perform the numerical simulations is available online in Dryad repository at

- Ref. [82]: Data from: Competing metabolic strategies in a multilevel selection model, André Amado, Lenin Fernández, Weini Huang, Fernando F. Ferreira, Paulo R. A. Campos, Dryad Digital Repository (2016), DOI: 10.5061/dryad.q6784

3 EFFECT OF TRADEOFFS IN CELL DIFFERENTIATION

Highlights

A model for the evolution of specialization under an arbitrary network of tradeoffs is introduced

The statistical properties of the model are studied

An application of the model to a concrete biological system (cyanobacteria) is developed

The concept of tradeoff is a central one in evolutionary biology. Tradeoffs are at the subject of a very active discussion in scientific community due to their contribution in shaping life histories and ecological/evolutionary dynamics in nature [86]. Tradeoffs introduce constraints in the evolution process by forcing living organisms to choose: they cannot optimize all traits simultaneously. This creates room for the evolution of specialization and diversification of life, generating complexity and interdependence in the ecosystems. As such, they are believed to play an essential role in creation and maintenance of diversity in life [87], as well as a pressure towards the establishment of the division of labor [88]. This issue is addressed in the context of several distinct frameworks, such as evolutionary game theory [89–92], resource-based modeling [79, 85, 93, 94], developmental plasticity [95, 96], and so on.

This chapter is dedicated to the concept of tradeoff itself and its consequences for the evolution of complexity. Tradeoffs reveal themselves as negative correlations between traits that prevent simultaneous optimization [97, 98]. These correlations lead to a situation where the living organisms are forced to choose between performing the several functions poorly or specializing in one of them.

Tradeoffs are usually depicted in literature as relations between pairs of variables. Very few theoretical studies take into consideration that tradeoffs appear in nature frequently as complex networks of interactions between traits, despite a growing body experimental evidence for this fact [99–105]. In our view, multidimensional tradeoffs are especially important in the study of the evolution of cell specialization in multicellular organisms. As such, we are interested in producing a model that can address this question.

3.1 Model

For the sake of an easier comprehension of the model, let us first introduce an overview of the same, with the details postponed to a later stage. The model consists of an asexually reproducing population of cells organized in colonies. We consider clonal development[1] from a unicellular propagule, giving rise to multicellular organisms (colonies) as cells undergo binary fission [96]. Clonal development keeps the genetic variation low among cells within the same organism since it fundamentally stems from somatic mutations[2]. This leads to reduced competition within the organism. It has long been thought that clonal life cycles replaced nonclonal ones, but recent research suggests the possibility that clonal development has been present since the beginning of multicellularity and has been preserved instead due to the significant advantage it provides [106]. In clonal development, each cell in a colony undergoes several rounds of binary fission, until a maximum size S is reached[3]. We assign a mutation probability μ per gene to each cell division, which corresponds to an effective mutation rate of $2S\mu$ per life cycle per gene[4]. When size S is attained, the colony is subjected to viability selection[5]. The viability selection takes

[1] In this context, clonal development is the development of a new organism from a single initial cell. It does not specify if reproduction is sexual or asexual. It is defined in contrast with aggregative development, where cells from different origins gather together to originate a new multicellular organism.

[2] A mutation that is not inherited from the parent and happens in cells with somatic function.

[3] Alternatively, we could express the size through the number of cell divisions necessary to achieve it, $\log_2(S)$.

[4] Starting from a single cell we need $\log_2 S$ binary divisions to achieve S cells. In step n there are 2^n cell divisions, therefore the total number of division per cycle is $\sum_{n=1}^{\log_2 S} 2^n = 2(S-1) \approx 2S$, for large S. Thus, the mutation rate per cycle is approximately $2S\mu$.

[5] The fitness of an organism can frequently be decomposed in a product of two main components: the viability, which corresponds to the probability of the organism to stay alive until reproduction age, and the fertility, which is the number of offspring an organism can have. We introduce two selection phases that directly match these two components.

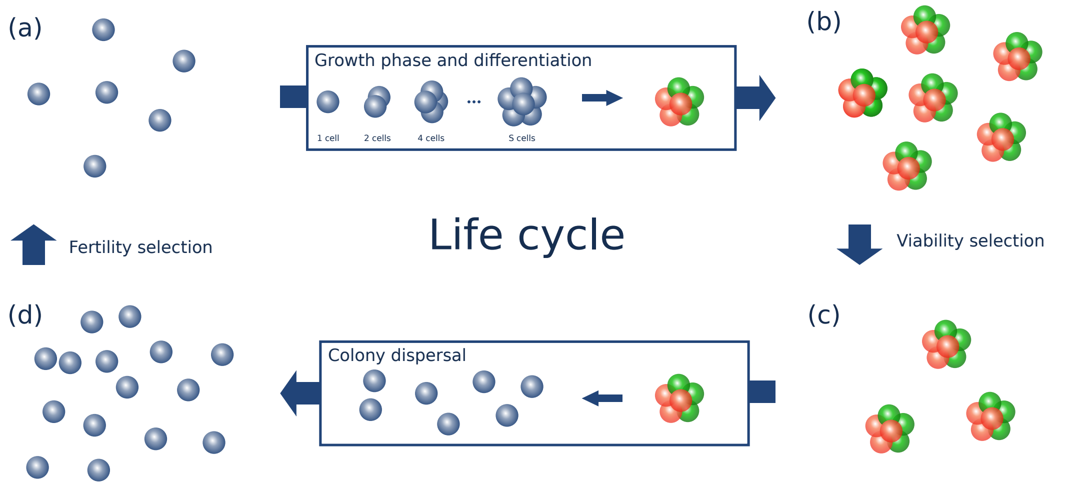

Figura 3.1: Illustration of the life cycle of the model. It starts with a stage (a) where the organisms are unicellular and undifferentiated, constituting a spore/propagule ready to found a new colony. Each propagule then undergoes cellular division up until the colony reaches size S and its cells differentiate according to the instructions contained in the regulatory genes. In each cellular division, each gene can mutate with probability μ. We have now an adult population (b). The elements of this population compete and the ones that survive viability selection reach reproductive age (c). The surviving colonies disperse completely producing a set of unicellular propagules (d). The propagules that endure fertility selection restart the cycle, producing new colonies. Figure adapted from [107].

into account the competition between the aggregates in a limited environment possessing a maximum carrying capacity K [6]. Then, the surviving colonies disperse, giving rise to unicellular propagules which will restart the process and form new colonies[7]. Not all cells are able to originate new colonies and the unicellular stage must endure another level of selection, the fertility selection. For the purposes of the model here proposed, we consider a fixed fertility f but we keep the fertility selection stage since a varying fertility may be an important ingredient in generalized versions of this work. The life cycle here described is illustrated in Fig. 3.1. A life cycle akin to this, with a dispersal phase and a growing phase from a unicellular propagule, has been recently observed in the lab to emerge in the transition to multicellularity from the previously unicellular alga *Chlamydomonas reinhardtii* by Ratcliff et al. [51]. In some systems, like the one described in Ref. [51], a dispersal phase very close to the one modelled here is realized. However, in most biological

[6] Carrying capacity of a system is maximum population the environment of the system can sustain. We refer to maximum carrying capacity K because in this work the carrying capacity is smaller or equal to K. Further details are given below.

[7] An alternative setting would be to have the groups split in two. This would lead to a higher genetic diversity within each group since the genetic bottleneck introduced by the unicellular stage did not exist anymore.

systems, the fact that only a fraction of the cells of the parent organism give rise to new individuals can be represented by a combination of a dispersal phase with fertility selection.

We now must delve into the model and life cycle implementation. The most relevant phase for our analysis of the life cycle is the viability selection since we will focus on tradeoffs between somatic functions[8]. For a colony to survive adulthood, it must compete with all the other colonies. The aggregates which achieve higher viability v have a larger probability of survival. Given the corresponding viability, an organism survives until reproduction age with probability

$$\left[1 + (S-1)\frac{N}{Kv}\right]^{-1}, \tag{3.1}$$

where N denotes the total number of colonies, S is the size of the colony immediately before the unicellular stage and K the maximum carrying capacity of the population. We here refer to the maximum carrying capacity of the system because, as the viability v changes, the carrying capacity is effectively altered, with the maximum carrying capacity K being achieved when all colonies have viability equal to one. A system populated by colonies with a low viability will support a smaller population compared to colonies with a high viability. This is a modified version of the Beverton-Holt stock-recruitment model which assumes that the per capita number of offspring is inversely proportional to a linearly increasing function of the number of mature colonies [108]. No adult colonies are transported between generations since they disperse in the transition to the unicellular stage. This way, there are no overlapping generations in the model.

The viability depends on the organism's performance of the tasks or biological functions it should accomplish to survive. As such, the exact form of the viability should reflect the underlying biology of the organism. We denote the performance of the organism at task i by φ_i. We will assume all functions to be essential. From the several possibilities available we consider the geometric mean of the φ_i-values. Other possibilities could include the harmonic mean or more complex functions of the φ_i-values. Note that these functions do not need to be symmetric in the

[8] Somatic functions are the support functions that keep the organism alive and are hence related to the viability of the organism. Besides somatic functions, there are germinative functions, which are responsible by the fertility of the organism.

arguments and this symmetry is here assumed for simplicity reasons. With this choice the viability takes the form

$$v = \sqrt[n]{\prod_i \varphi_i}\,, \tag{3.2}$$

where n stands for the number of biological tasks to be carried out. For comparison, in Ref. [96] a system is discussed where a single somatic function determines the viability.

As the biological tasks under scrutiny present tradeoffs among them, increasing the output of one task may have a detrimental effect on others. We introduce the tradeoff considering that the output of a certain task i is codified by a structural gene Y_i. This gene, in our implementation, can assume continuous values in the range $[0, 1]$ and represents the investment of the cell in the corresponding task i. Taking this into account, a simple way of representing the tradeoffs is to assign the performance φ_i of task i to

$$\varphi_i = Y_i^{T_{ii}} \prod_{j \neq i} [1 - Y_j]^{T_{ij}}\,. \tag{3.3}$$

Thus, the tradeoffs are encoded in the matrix $\mathbf{T} = \{T_{ij}\}$. If this matrix is diagonal no tradeoffs are present since the output of function i reduces to $\varphi_i = Y_i^{T_{ii}}$ and receives no penalty from the remaining functions. On the other hand, nondiagonal elements of $\mathbf{T}$ introduce a decrease in the performance of other functions, for instance, if $T_{ij} \neq 0$ the output of function i is atenuated by a multiplicative factor of $[1 - Y_j]^{T_{ij}}$. Therefore, each gene Y_i has two effects: a direct effect increasing the performance of task i and an indirect effect decreasing the output of the remaining tasks for which T_{ji} is not null. This way, Eq. 3.3 captures the essence of the tradeoff concept. If the organism is to increase its fitness, it is required to find ways of offsetting this effect. One important way available in multicellular organisms' toolset to deal with this is to segregate incompatible functions into different cells, thus mitigating the cellular level tradeoff. Though, this does not come without costs since it involves regulatory genes and different responses to intercellular chemical signalling. Nevertheless, differentiated response to complex environmental and within-colony chemical signalling is already something that all organisms are prepared to, including unicellular organisms [109]. To take regulation into consideration we must modify the performance function in Eq. 3.3 in order to include differentiated responses to each stimulus, leading to nonuniform behavior across the cells of the aggregate. With this in mind, we introduce a set of regulatory genes $\{y_{ik}\}$. The gene y_{ik} suppresses the activity of the structural gene Y_i in a cell subject to stimulus k. Therefore, the contribution of a cell subject to chemical signal k to the overall somatic function i, φ_{ik}, of the colony can be codified as

$$\varphi_{ik} = [(1 - y_{ik})Y_i]^{T_{ii}}\, c(y_{ik}) \prod_{j \neq i} [1 - (1 - y_{jk})Y_j]^{T_{ij}}\, c(y_{jk}), \tag{3.4}$$

where $c(y_{jk})$ is a cost function, explained below. After some algebra, this equation can be equivalently rewritten in a more compact form, as

$$\varphi_{ik} = \prod_{j} |1 - \delta_{ij} - (1 - y_{jk})Y_j|^{T_{ij}}\, c(y_{jk}). \tag{3.5}$$

The genes y_{ik} introduce the possibility of blocking the function i in a cell under stimulus k, eventually reducing the effect of the tradeoff by segregating incompatible functions to different cells. If a cell is under stimulus k and all of its regulatory genes y_{ik} are close to zero, it will be capable of performing all functions since no structural gene Y_i is suppressed. On the other hand, a cell that has some of its regulatory genes y_{ik} close to one will not undertake the corresponding functions and so it must coexist with other cells within the same organism that can perform those suppressed functions, as all functions are said to be essential. As the system evolves, incompatible cellular processes tend to suppress the expression of genes encoding other functions [110], thus contributing to the formation of aggregates with permanently specialized cellular functions. The mechanisms of gene suppression and developmental plasticity embody a cost in fitness terms which is incorporated into the estimation of φ_{ik} through the cost function $c(y_{ik})$, as a decreasing function of the regulation effect y_{jk}. This means that the stronger the suppression is, the more costly it becomes [96, 111]. In the present implementation, a Gaussian function $c(y) = \exp(-\frac{1}{2}\frac{y^2}{\sigma_y^2})$ is considered as cost function.

Due to regulation, the cells have different contributions to the output of each function. Therefore, we should consider average contribution over the groups of cells subject to each stimulus k [88], i.e.,

$$\varphi_i = \frac{1}{S} \sum_k \varphi_{ik}, \tag{3.6}$$

where we recall that S stands for the number of cells in each organism.

This model is partially inspired by the one introduced by Gavrilets in 2010 [96], where the author models the development of division of labor via evolution of developmental plasticity. It shows that division of labor could arise fairly quickly from simple undifferentiated organisms. The model proposed by the author deals with one tradeoff between somatic and reproductive functions only. Our approach extends the work done allowing one to probe an arbitrarily complex network of tradeoffs and study the effect on the evolution of cell differentiation.

As a reference guide to the reader, the parameters of the model are summarized in the box below.

Notation summary

s **(tradeoff strength):** In the simplest case, the tradeoff strength, s, is uniform over all the tradeoffs. Therefore, under the assumption of a uniform tradeoff strength T_{ij} is either equal to zero (if there is no tradeoff between a given pair of genes) or s. The assumption of a uniform tradeoff strength will be released later. In such a situation, the strength s is not a constant but rather taken from a given probability distribution.

v **(viability):** The viability is a measure of how adapted the organism is to its environment. It determines the probability of the survival of the organism until reproduction age.

f **(fertility):** After surviving viability selection, each cell of the colony can give rise to a newly formed colony with probability f, the fertility of the cell.

μ **(mutation probability):** During cell division, there exists a uniform probability of mutation per gene, μ. If a mutation takes place in a given gene j, Y_j (in case it is a structural gene) or y_{jk} (in case it is a regulatory gene) changes to a randomly chosen value from a uniform distribution $[0, 1)$.

K **(maximum carrying capacity):** The maximum carrying capacity, K, corresponds to the population size upon maximum fertility, $f = 1$ (all cells can successfully establish a new colony), and maximum viability, $v = 1$.

n **(number of tasks):** Number of biological functions or tasks to the performed by the organism.

t **(number of tradeoffs):** There are up to $n(n-1)$ tradeoffs, the number of degrees of freedom of the T_{ij} matrix.

3.2 Simulation Protocol

We start the simulation with a population of $N_0 = 100$ genetically uniform colonies at the unicellular stage. The initial cells are ascribed a genotype of $Y_i = 0.75$ and $y_{ik} = 0$, which corresponds to a population of colonies that do not possess cell differentiation. The colonies grow until reaching the adult size of S. During this growth stage the cells can suffer mutation according to the rules previously introduced, which correspond to an effective mutation rate of $2S\mu$ per colony per gene in each cycle. For instance, values of $S = 16$ and $\mu = 10^{-5}$ yield an effective mutation rate of 3.2×10^{-4}. We consider that the gene mutates to a uniform value in the range $[0, 1)$. While this can include large mutations, the effect on the phenotype should be relatively small since each organism has many genes and only one is affected per mutation. When the adult size S is achieved, the aggregates undergo viability selection following Eq. 3.1, using the viability values calculated according to Eqs. 3.2 to 3.6. The environment is assigned a maximum carrying capacity K, which is taken to be $50,000$ in our simulations unless stated otherwise. The colonies that survive viability selection then disperse, giving rise to new unicellular propagules. Each of these propagules originates a new colony with probability given by its fertility f or dies otherwise. In our implementation the fertility is taken to be constant since we are focusing on somatic tradeoffs, but this is not the case in general. See the work by Gavrilets [96] for an example of a study that considers a dynamical fertility rate, based on a tradeoff between one somatic and one germinative task. The surviving propagules are taken as the founding population of the next generation. This process is repeated until a stationary population is achieved. Typically, the system is allowed to evolve for 5×10^6 generations, after which measurements are taken in the following 5×10^6 generations.

This process is repeated a number of times, 1000 in our implementation unless stated otherwise. The results of these runs are used to produce statistical averages. As we are interested in the general properties of the tradeoffs, as opposed to a specific system, each run uses a given random realization of the matrix T_{ij}, subject to some constraints, such as specific number of tradeoffs and tradeoff strength.

A criterion needs to be established as a measure of the differentiation level among cells. We adopt as metric the distance d_{ij} between the response to two stimuli i and j as

$$d_{ij} = \sqrt{\frac{\sum_{k=1}^{n} (y_{ki} - y_{kj})^2}{n}}. \tag{3.7}$$

If the distance d_{ij} is higher than a critical value d_c we consider that a differentiated response to those stimuli is present, meaning that the cells differentiated into specialized types. Since $y_{ij} \in [0, 1]$, d_{ij} also lies in the range between 0 and 1. Unless stated otherwise, the threshold d_c is set at $d_c = 0.2$, which provides a solid criterion for determining the differentiation among the cells, as will be shown and discussed later, in Sec. 3.5.

As previously mentioned, the tradeoff strength between each pair of functions is described within the context of the present model by the tradeoff matrix T_{ij}

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1M} \\ T_{21} & T_{22} & \cdots & T_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ T_{M1} & T_{M2} & \cdots & T_{MM} \end{bmatrix}. \tag{3.8}$$

The diagonal element $\mathbf{T}_{ii}$ defines the output of task i as a function of Y_i, in the absence of tradeoffs. On the other hand, the off-diagonal element $\mathbf{T}_{ij}$ $(i \neq j)$ quantifies the intensity of the tradeoff interaction between task i and task j. We have t tradeoffs, where t can takes integer values from 0 to $n(n-1)$, corresponding to the number of nonzero off-diagonal entries. In the simulations, we will analyze two cases: tradeoffs of constant strength and strength drawn from a given probability distribution.

3.3 Analytical estimation of the number of colonies

A full analytical treatment of the system is rather difficult but some results can be obtained, such as the estimation of the number of colonies at equilibrium. Analytical results are always important since they can provide some insight on the model and help to analyze the simulations.

We define the fitness W of an individual as the product of the survival probability until adult age with the expected number of offspring an adult can produce. This is in line with the usual definition of fitness in the context of the study of the evolution of life history traits. Thus, we have

$$W(N) = \frac{1}{1 + (S-1)\frac{N}{Kv}} \times fS. \tag{3.9}$$

Each successful offspring will establish a new colony at the next generation. In a discrete-time model, the population at the next generation is therefore given by

$$N_{t+1} = G(N_t) = W(N_t)N_t = \frac{fS}{1+(S-1)\frac{N_t}{Kv}}N_t. \tag{3.10}$$

To find the equilibrium population we require that $N_{t+1} = N_t = \hat{N}$, i.e., a stationary condition. Under this condition, Eq. 3.10 has two equilibria

$$\hat{N}_0 = 0 \qquad \text{and} \qquad \hat{N}_1 = \frac{fS-1}{S-1}Kv. \tag{3.11}$$

These equilibria correspond, respectively, to the extinction of the population and a stationary population. The equilibria are guaranteed to be stable if $-1 < \left.\frac{\partial G}{\partial N_t}\right|_{N_t=\hat{N}} < 1$ (check appendix A). Calculating the derivative of the Eq. 3.10 one obtains

$$\begin{aligned} \frac{\partial G}{\partial N} &= \frac{fS}{1+(S-1)\frac{N}{Kv}} - \frac{fS}{\left[1+(S-1)\frac{N}{Kv}\right]^2}\frac{N(S-1)}{Kv} \\ &= \frac{fS}{1+(S-1)\frac{N}{Kv}}\left[1-\frac{N(S-1)}{Kv+(S-1)N}\right]. \end{aligned} \tag{3.12}$$

Therefore, the zero solution is stable if $-1 < fS < 1$. As fS is always positive only the second condition needs to be guaranteed and we get $fS < 1$. The condition obtained for the second solution is somewhat more complex. Its stability is dictated by the condition

$$\begin{aligned} -1 < \left.\frac{\partial G}{\partial N}\right|_{N=\hat{N}_1} < -1 \Rightarrow -1 < \frac{fS}{1+(S-1)\frac{\frac{fS-1}{S-1}Kv}{Kv}}\left[1-\frac{\frac{fS-1}{S-1}Kv(S-1)}{Kv+(S-1)\frac{fS-1}{S-1}Kv}\right] < -1 \\ \Rightarrow -1 < \frac{fS-(fS-1)}{fS} < 1 \quad \Rightarrow \quad -1 < \frac{1}{fS} < 1 \quad \Rightarrow \quad -1 > fS > 1. \end{aligned} \tag{3.13}$$

Once more the first inequality is always verified and we are left with the condition $fS > 1$. This is an intuitive result since fS represents average number of daughter colonies that each colony will produce. If each colony produces less than one offspring in average obviously the population will go extinct. In this analysis the population is continuous and therefore it can achieve an arbitrarily small size at equilibrium. For this reason, the stability of the solution does not depend neither on K nor v since these quantities mainly determine the equilibrium population size. In the full discrete system, a dependence on the product Kv is expectable. A small Kv implies a small equilibrium population, that can easily be extinguished due to the fluctuations induced by the stochastic nature of the system. This also entails that the system is more susceptible to extinction in the early stages of the evolutionary dynamics, when division of labor has not evolved yet and the vibility of the aggregates is low.

For a typical set of parameters, if $v \approx 1$ (no tradeoff), the equilibrium population yields $\hat{N} = \frac{0.5\times16-1}{16-1}\ 5 \times 10^4 \approx 2.3 \times 10^4$, which perfectly matches the equilibrium population sizes found in the simulations. When strong tradeoffs are present and the population overcomes extinction, it evolves to an equilibrium situation where the tradeoffs are mitigated through division of labor. Even after specialization takes place, the existence of tradeoffs reduces the population at equilibrium since the specialization leads to less cells performing each task. Also, as one can see in Eq. 3.4, the contribution of the specialized cells picks up factors of $\exp(-1/2\sigma_y^2)$ due to the regulation costs.

3.4 One tradeoff case

Let us start by analyzing the simplest case of one tradeoff, before moving on to the general situation. It turns out that it is possible to provide good analytical approximations for the viability in this case (for detailed calculations, please refer to Appendix D). We will consider a tradeoff matrix with diagonal elements $\mathbf{T}_{ii} = s$ and all off-diagonal elements null except one. This nonzero element has value s^*. In the case when none of the cells specialize and the aggregate is totally generalist, the viability of such a system takes the form

$$\overline{v} = \left[\frac{s^s s^{*s^*}}{\left(s+s^*\right)^{s+s^*}}\right]^{1/n} . \tag{3.14}$$

The other limiting case is when the cells subject to tradeoffs specialize completely. In this limit, the viability can be found to be given by

$$v = \frac{1}{n}\left[(n-1)\,c(1)\,[n-1+c(1)]^{n-2}\right]^{1/n} . \tag{3.15}$$

Notice that, in the case of Eq. 3.14, the viability is a decreasing function of the tradeoff strength. Conversely, in Eq. 3.15, the viability becomes independent of the tradeoff strength due to the total specialization of the cells.

Fig. 3.2 depicts the maximum viability achievable in the cases of all cells being generalists, fully specialized or partially specialized in such a way that fitness is

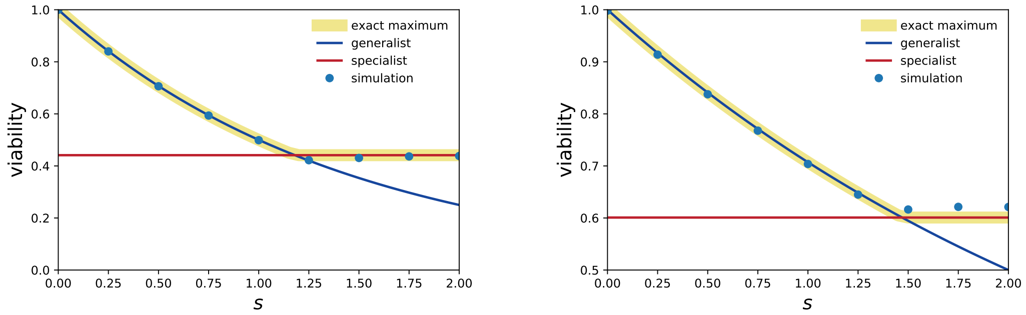

Figura 3.2: Maximum viability obtained by a colony as a function of the tradeoff strength s (s^* also set to s) for the numerical maximum (yellow line), all cells generalist (blue line), total specialization (red line) and simulation results (circles). The left panel shows the result for $n = 2$ functions and the right panel for $n = 4$ functions. The remaining parameter, σ^2, is 2.

maximized. The partially specialized case is obtained from the numerical study of the full expression. We can see that in practice the numerical result is very close to either totally generalist cells or fully specialized cells, whichever provides the highest fitness. The transition between these two regimes happens very abruptly. In the simulation results, we can see that the evolutionary mechanisms are very efficient selecting for the optimum behavior, the result is very close to the theoretical maximum a population can reach. Near the transition between regimes, the simulation result is slightly lower than the maximum achievable. This happens because the viabilities for both generalist and specialized aggregates are very close leading to a weaker selection. The system is almost in a neutral selection regime. As the simulation is initiated with only generalist aggregates, the fixation time of a specialist mutant is very large.

Fig. 3.3 provides an example of the adaptation process in time. As time advances, beneficial mutations are fixed in the population, progressively increasing the population fitness until a configuration close to the maximum is found. Even after a very long time the genes have some oscillations, which correspond to alternation between slightly different strains with fitness close to the maximum. In any finite population, a new variant with very close fitness can get fixed in the population through neutral drift even if it is slightly deleterious.

3.5 General case

Here we analyze the situation in which tradeoffs of constant strength are in place. Analytical calculations with more than one tradeoff are of course possible but soon

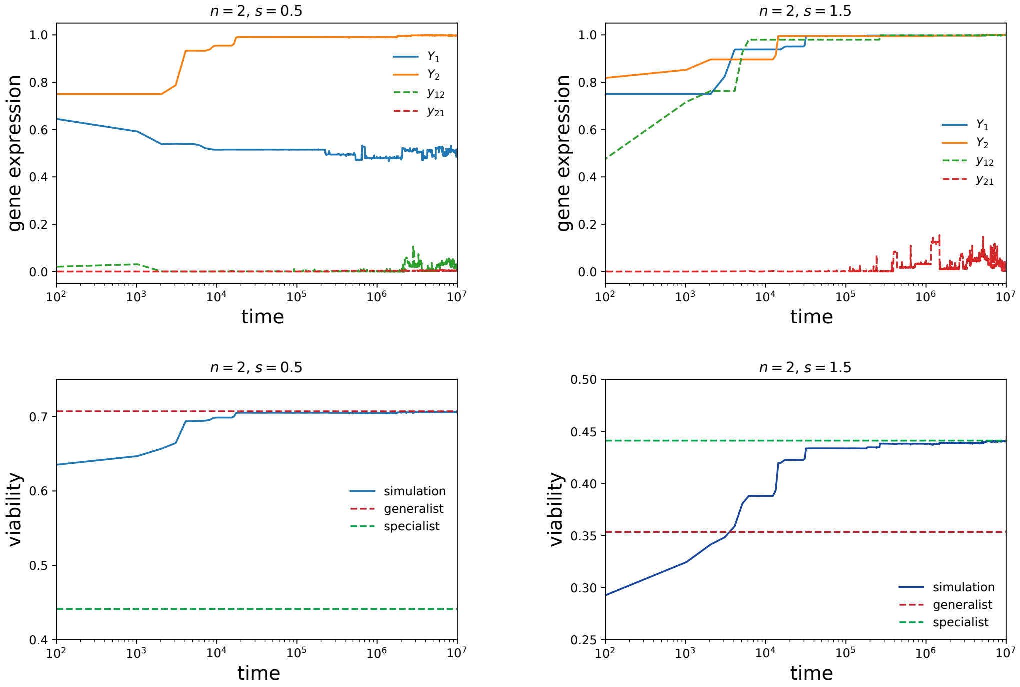

Figura 3.3: Top panels: time evolution of the gene expressions for tradeoff strength $s = 0.5$ (left panel) and $s = 1.5$ (right panel), for 2 functions. Bottom panels: time evolution of the corresponding viabilities. The viabilities of a generalist (red) and specialized (green) organism are shown in dashed lines. The remaining parameter, σ^2, is 2.

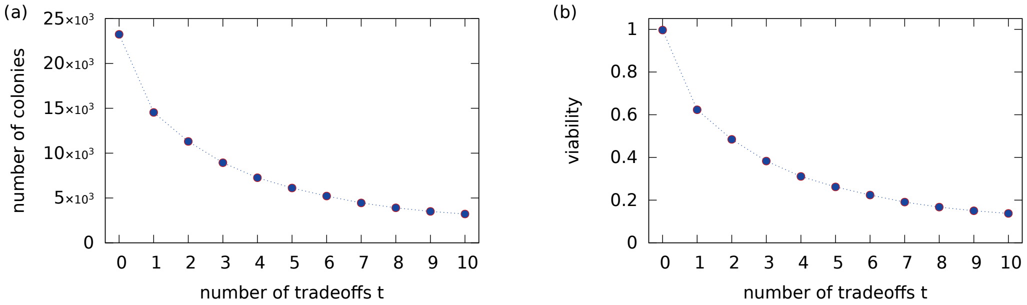

Figura 3.4: Panel a: Average number of colonies versus the number of tradeoffs t for four biological functions (the remaining parameters are $s = 2$, $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Panel b: Average viability of a colony versus the number of tradeoffs t for four biological functions (the remaining parameters are $s = 2$, $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Each point is an average over 1000 independent configurations. Figure adapted from [107].

become too cumbersome to be presented here. Unless stated otherwise it is assumed that number of cells before the unicellular stage is $S = 16$, mutation probability $\mu = 10^{-5}$, fertility $f = 0.5$, tradeoff strength $s = 2$ and carrying capacity $K = 50\,000$. The initial number of colonies is 100. As previously stated, we initiate the simulations with all the structural genes set to $Y_i = 0.75$ and the regulatory genes to $y_{ij} = 0$. In this situation, the cells are in an undifferentiated state, where all the functions are active in each cell.

In Fig. 3.4a, one can see that the number of colonies decreases with the number of tradeoffs, as expected, reflecting the extra costs introduced by the specialization. In the absence of tradeoffs, the mean viability goes to one, meaning that all traits can be maximized simultaneously as there are no constraints. As tradeoffs are added, specialization requires the suppression of the expression of an increasing number of genes entailing a greater cost in terms of fitness.

Fig. 3.5a explores the effect of the tradeoff strength s on the evolution of the system, now for several values of the number of tradeoffs t. We can see that, similarly to the situation with only one tradeoff, there is a threshold between an initial region of tradeoff strength where the aggregates adopt a generalist configuration and another region where a specialized configuration is preferred. In the first region, the population decreases because the generalist configuration loses viability with the tradeoff strength. As soon as the aggregates specialize, a plateau is achieved for the population size since the specialized cells are not directly affected by the tradeoff anymore. There is a point beyond which the population cannot be sustained even under total specialization and goes extinct. Fig. 3.5b shows the probability of extinction of the population during the timespan of the simulation. One can see

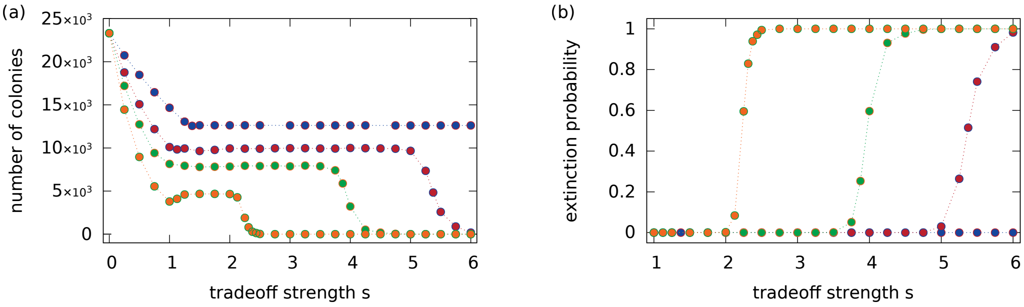

Figura 3.5: Panel a: Average number of colonies versus the strength of the tradeoff s for three biological functions and one (dark blue), two (red), three (green) and six (orange) tradeoffs (the remaining parameters are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Panel b: Extinction probability versus the strength of the tradeoff s for three biological functions and one (dark blue), two (red), three (green) and six (orange) tradeoffs (the remaining parameters are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Each point is an average over 1000 independent configurations. Figure adapted from [107].

that there is a steep transition between a regime without extinction to a regime where the population is doomed to extinction. The higher the number of tradeoffs, the sooner the extinction regime comes into play. As the number of tradeoffs is increased, the transition becomes sharper.

For the sake of completeness, we also survey the dependence of the number of colonies on the size of group just before the unicellular state S and the maximum carrying capacity K. These two parameters impact the survival probability of a colony, as can be inferred from Eq. 3.1. We show the average population sizes as a function of both quantities in Fig. 3.6, for the case of three biological functions and three and six tradeoffs. The average population size displays an interesting behavior when S is varied. After an initial growth for small S, the population saturates and then falls. Notice that the behavior is the same for three and six tradeoffs, although for three tradeoffs the population decrease starts at a much larger S. This decrease in the population size is due to a much larger extinction probability. This outcome shows that the colony size cannot be enlarged without bound in this life cycle as its increase in size leads to a reduced survival probability.

As expected, in the limit of large K, the continuous model and the discrete simulations match and we have a linear dependence of the population size on K. This is evinced by the linear fit shown as a blue line in Fig. 3.6b. Also, there is a minimum carrying capacity below which the population is doomed to extinction due to the stochastic fluctuations in the population size. Having a low carrying capacity has a similar effect on the population dynamics as a small viability.

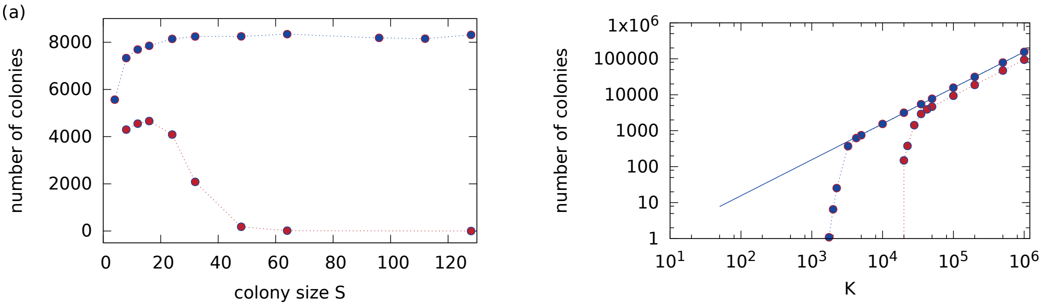

Figura 3.6: Panel a: Average number of colonies against its size just before the reproduction stage (S), for three biological functions and three (blue) and six (red) tradeoffs (the remaining parameters are $s = 2$, $\mu = 10^{-5}$ and $K = 50\,000$). Panel b: Average number of colonies versus the maximum carrying capacity of the system (K) for three biological functions and three (blue) and six (red) tradeoffs (the remaining parameters are $s = 2$, $S = 16$ and $\mu = 10^{-5}$). The straight line is a linear fit which matches the data well, as expected, in the limit of large K (please see Eq. 3.11). Each point is an average over 1000 independent configurations. Figure adapted from [107].

However, if the number of tradeoffs is enlarged ($t = 6$ in the plot) we already observe an abrupt drop of the number of colonies at intermediate S, due to the extinction of the population. Indeed, the fall in the number of colonies is also found for $t = 3$ but this effect occurs at much larger S. This outcome also shows that the colony size S can not be enlarged without bound as its increase in size reduces the probability of survival. This critical colony size depends on the number of tradeoffs t.

Having characterized the population sizes, we want now to focus our attention on the differentiation level attained within the aggregate. The number of differentiated types is the number of different cell phenotypes arising as a response to the different stimuli to which the cells are subject in the organism. In our model, the maximum number of different cell types is equal to the number of biological functions or tasks an organism must perform. The information on the cell differentiation is encoded in the regulatory genes y_{ik}. With this in mind, in Sec. 3.2, we have introduced a metric to characterize the amount of differentiation between cell types in an organism. Fig. 3.7 shows the response of our results to the criteria adopted for d_c. For very low values of d_c, the definition flags any small fluctuation in the regulation genes y_{ik} as a new cell type, which is not biologically realistic. At the other end of the spectrum, if $d_c \approx 1$ all the cells are characterized as of a same type, even if they present significant differences. At last, another plateau exists in-between for which an intermediate value of the number of cell types is stable. As this number remains constant for a large range of d_c, roughly $d_c \in [0.003 : 0.7]$, this

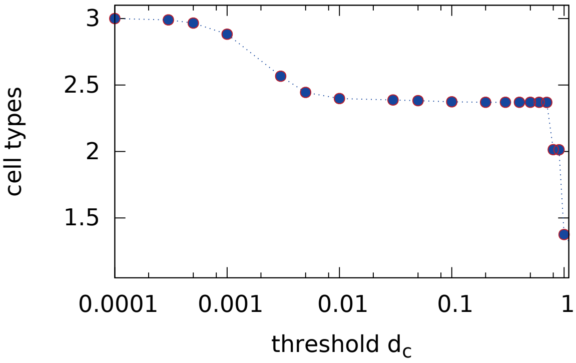

Figura 3.7: Average number of cell types as a function of the threshold distance d_c for three biological functions and three tradeoffs. The remaining parameters are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$. Each point is an average over 1000 independent configurations. Figure adapted from [107].

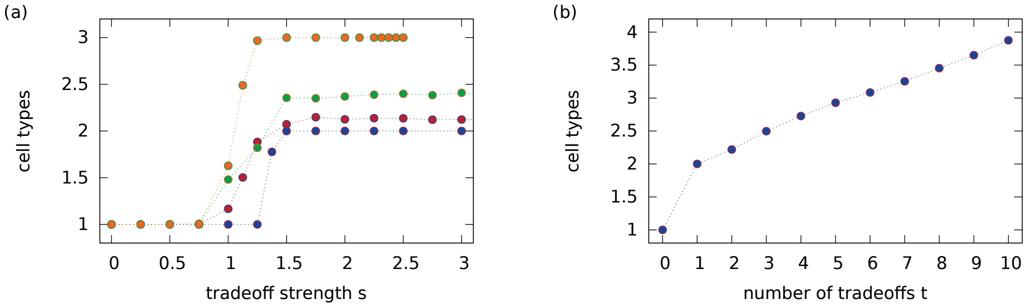

Figura 3.8: Panel a: Average number of cell types versus the strength of the tradeoff s for three biological functions and one (dark blue), two (red), three (green) and six (orange) tradeoffs (the remaining parameters are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Panel b: Average number of cell types versus the number of tradeoffs t for four biological functions (the remaining parameters are $s = 2$, $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$). Each point is an average over 1000 independent configurations. Figure adapted from [107].

method actually provides a robust characterization of the cell differentition in an aggregate. According to what was mentioned before, we adopt the value $d_c = 0.2$ to characterize the number of independent cell types in the aggregate.

Figure 3.8a shows the average number of cell types obtained for different tradeoff strengths, in the case of three biological functions. We can see that, independently of the number of tradeoffs involved, the number of different cell types grows with the tradeoff strength until it reaches a plateau. The ultimate value achieved by the plateau depends on the number of tradeoffs involved. As the number of tradeoffs rises, the minimum strength necessary to cause differentiation decreases. On the other hand, Fig. 3.8b displays the dependence of the number of cell

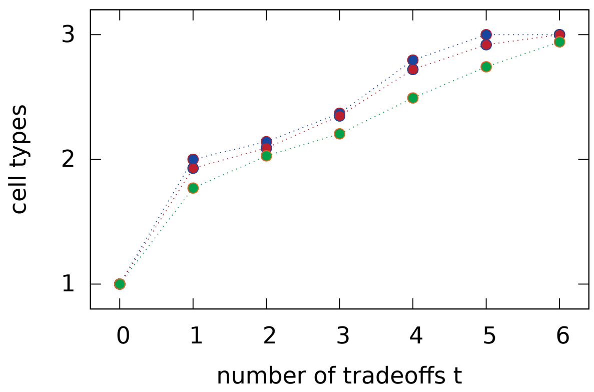

Figura 3.9: Average number of cell types as a function of the number of tradeoffs t for three biological functions. Here the tradeoff strength is variable and drawn from an uniform distribution. The blue points correspond to a constant tradeoff $s = 2$, the red points denote an uniform distribution with $s \in [1, 3]$, whereas the green points also denote an uniform distribution with $s \in [0.5, 3.5]$. The remaining parameters are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$. Each point is an average over 1000 independent configurations. Figure adapted from [107].

types with the number of tradeoffs, while keeping the tradeoff strength constant. It is clear that an increase in tradeoffs promotes cell differentiation. This can be achieved by either increasing the number of tradeoffs, i.e. the number of non-null elements of the tradeoff matrix T_{ij}, or increasing the strength of those tradeoffs. For four functions, as shown in Fig. 3.8b, the maximum specilization is achieved for 10 tradeoffs.

Variable tradeoff strength

Here we release the assumption of constant tradeoff strengths. The tradeoff strength s is now drawn from a uniform distribution $s \in [s_{inf}, s_{sup}]$. The tradeoff strength varies not only among different Monte Carlo runs but also across different pairs of genes. To facilitate the comparison with the previous result, we consider s_{inf} and s_{sup} in such a way that their average is kept at two, i.e. $\langle s \rangle = 2$. Therefore, the only difference to the previous case is the variance of the distribution, which was zero in Fig. 3.8b. The result can be seen in Fig. 3.9. One can see that for larger variance on the tradeoff strength the specialization decreases slightly, although the general qualitative behavior remains unaltered.

We will analyze the model in a generic form, performing a statistical study over the possible configurations given certain restrictions on the number of tradeoffs and/or their strength. Nevertheless, modifications of this model can be applied to specific systems in nature. For clarity and increased biological intuition, we introduce an example in which the model can be applied. In the following, the model is used as a simplified version of cyanobacteria.

3.6 Concrete example: cyanobacteria system

Until this point we analyzed generic systems, where the tradeoffs were not attached any concrete biological meaning. In this subsection we study a concrete system: the cyanobacteria. Cyanobacteria need both carbon compounds and nitrogen compounds [112]. For the carbon compounds they rely on photosynthesis, while the nitrogen can either be obtained from the environment, if available, or through nitrogen fixation [112]. The gas form of nitrogen is widely available since it is the main gas in the atmosphere. Nevertheless, it cannot be directly used by living organisms. It needs to undergo a process of nitrogen fixation, where it is incorporated in organic molecules and thus usable by organisms. The problem with this process is that the enzymes reponsible for nitrogen fixation are sensitive to oxygen, basically rendering photosynthesis and nitrogen fixation incompatible processes since carbon fixation releases oxygen [49, 112, 113]. This problem can be approached through several angles. Some species segregate these functions in time, alternating carbon and nitrogen fixation [113]. The downside of this approach is that it involves replacing a significant part of the enzymatic system periodically, which has significant costs, besides not allowing simultaneous carbon and nitrogen fixation. A more efficient solution has been developed in some species, evolving multicellularity and separation of labor, thus overcoming this restriction [49,112,113]. Some cells, denominated heterocysts, differentiate terminally and specialize in nitrogen fixation [49, 113]. The multicellularity in cyanobacteria evolved several times in history and, as far as it is known, represents one of the earliest forms of multicellularity, developed at least two billion years ago [114].

This problem fits nicely in our approach. The system can be described as having two required functions: carbon and nitrogen fixation. Nitrogen fixation is incompatible with carbon fixation, although the opposite is not true, so the tradeoff matrix can be represented as

$$\mathbf{T} = \begin{array}{c c} & \begin{array}{cc} C & N \end{array} \\ \begin{array}{c} C \\ N \end{array} & \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix} \end{array}, \tag{3.16}$$

where the letters C and N denote, respectively, carbon and nitrogen fixation. The off-diagonal 2 represents the incompatibility of nitrogen fixation with carbon fixation. As we are performing a qualitative analysis another value could have been taken, given that it is large enough to produce a strong tradeoff. We can also generalize the expression of the viability in Eq. 3.2 since the nitrogen can be provided either by the activity of the cell or supplied by an external source of biologically usable nitrogen

$$v = \sqrt{\varphi_C \left(\varphi_N + n\right)}. \tag{3.17}$$

Both carbon and nitrogen compounds are essential, but the nitrogen can be provided either by nitrogen fixation in the cell φ_N or, if available, by an external source n. Under these assumptions we aim to make a faithful reproduction of the scenario observed in nature within the perspective of the present modeling. We can now examine the effect of different concentrations of external biologically usable nitrogen, n, on the system.

This can be studied both analytically and through simulations. Adapting Eq. D.10 to this situation is straightforward and yields

$$\overline{v} = \sqrt{\frac{Y_C^2 \left[1 + \left(1 - y_{CN}\right)^2 c(y_{CN})\right]}{2} \left[\frac{\left(1 - Y_C\right)^2 + \left(1 - (1 - y_{CN})Y_C\right)^2 c(y_{CN})}{2} + n\right]} \tag{3.18}$$

Here Y_C and Y_N are the main genes associated to the fixation of carbon and nitrogen, respectively, and y_{CN} and y_{NC} are the corresponding regulatory genes. This is the result for a generic case, and we can obtain the limits of total specialization

$$\overline{v}_{spec} = \sqrt{\frac{1}{2}\left[\frac{c(1)}{2} + n\right]}, \tag{3.19}$$

and generalist

$$\overline{v}_{gen} = \sqrt{Y_C^2 \left[\left(1 - Y_C\right)^2 + n\right]}, \tag{3.20}$$

which, after optimizing for Y_C, becomes

$$\overline{v}_{gen} = \begin{cases} \frac{\left(3 - \sqrt{1 - 8n}\right)}{4} \sqrt{\left[\frac{1}{4}\left(\sqrt{1 - 8n} - 3\right) + 1\right]^2 + n}, & n \leq \frac{1}{2}\left(5\sqrt{5} - 11\right) \\ \sqrt{n}, & \text{otherwise.} \end{cases} \tag{3.21}$$

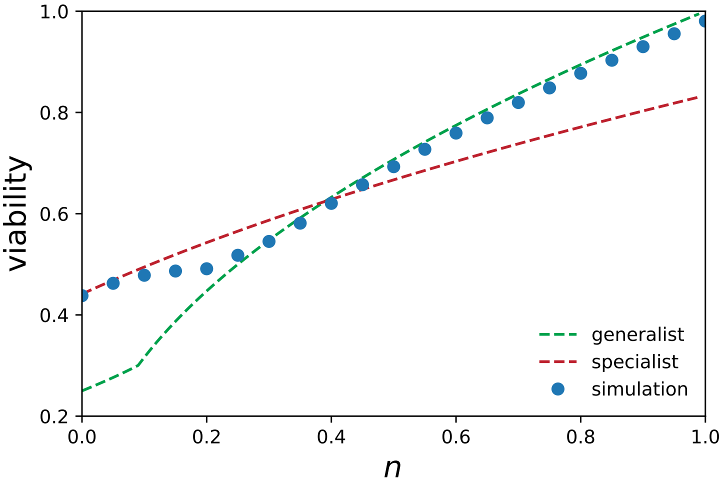

Figura 3.10: Viability as a function of the external nitrogen supply n. The blue dots correspond to the simulation results, the green dashed line to the theoretical generalist maximum and the red dashed line to the theoretical maximum under full specialization.

Comparing the expressions, one finds that the generalist solution acquires a higher fitness than the specialist when $n > \frac{c(1)}{2} \approx 0.389$. Let us now turn to the simulation results. Figure 3.10 shows the simulational and analytical results for the viability as a function of the external nitrogen supply n. We see that the simulation results start quite close to the result for full specialization, when the external nitrogen supply is very low. As the external supply is increased the result departs from the full specialization line and moves toward the generalist line. These results are confirmed when we analyze the average number of cells in Fig. 3.11a. One can see that for very low n we have always total differentiation. The degree of specialization decreases with n until only one cell type is present, at around $n \approx 0.4$. It is interesting to note that this is the point where the fitness of an organism composed of only generalist cells surpasses the fitness of an organism whose cells undergo total differentiation. The population size, shown in Fig. 3.11b, reveals a picture similar to the already analyzed viability. As expected, the population grows monotonically with the external nitrogen supply since it provides benefit without the costs of nitrogen fixation. The model's outcomes agree with the observed behavior of cyanobacteria response to nitrogen supply.

The model here described should be regarded as a toy model when applied to the study of cyanobacteria, for illustration purposes only, but we believe that introducing a simpler model improves the clarity. A more complex model could be considered within this framework, for instance by taking into account the tradeoffs

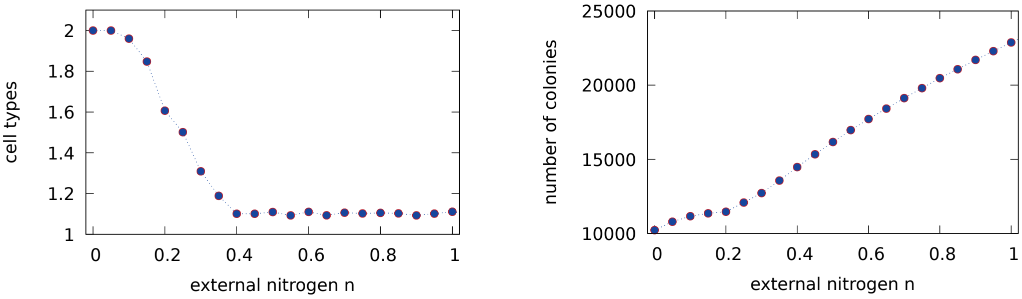

Figura 3.11: Panel a: Average number of cell types versus the concentration of external biological nitrogen. The remaining parameters of the model are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$. Panel b: Average number of colonies versus the concentration of external biological nitrogen. The remaining parameters of the model are $S = 16$, $\mu = 10^{-5}$ and $K = 50\,000$. Each point is an average of 1000 independent configurations. Figure adapted from [107].

between the currently described functions and reproductive function, which would lead to a more faithful representation of the system. Also, more general situations could be studied, for example, the equilibrium of the system when the external nitrogen is not fixed but variable on time, or the effect of letting the fraction of cells dedicated to each function dynamically evolve also.

3.7 Conclusions

We considered a system with tradeoffs between cell functions and analyzed the way these tradeoffs contribute to cell differentiation and division of labor. The cells are endowed with the possibility of suppressing the action of certain structural genes, as regulated by the regulatory genes upon given stimuli. This makes developmental plasticity possible. Of course developmental plasticity entails costs since it needs a complex network of regulatory genes and interactions. The system is adaptative and both structural and regulatory change through random mutation. Since we considered a constant fertility, our analysis concerns only somatic functions. The study is performed under different scenarios for the distribution of tradeoffs. We start the simulations in a situation where all the cells are completely undifferentiated, undertaking all functions regardless of the chemical stimuli the are subject to. As evolution proceeds they can suppress their contributions to some of the functions and mostly contribute to one or few tasks through the activation of regulatory genes that can suppress some of their activities when exposed to a given chemical stimulus. Although beneficial from the group perspective, the suppression mechanism produces a cost at the individual level.

The model shows that the tradeoffs affect not only the outcome of the division of labor but also the viability of the population as a whole. We have found that as the number of tradeoffs and tradeoffs strengths is increased the probability evolving differentiation increases too. The viability of the population decreases with the increase of tradeoff strength up to the point that differentiation compensates, since the tradeoff strength does not have an effect on the differentiated system anymore. Nevertheless, if the number of tradeoffs the differentiation costs grow significantly and can lead to an extinction of the population even in the case of total differentiation. Also, when the tradeoff strength is high, the population can be doomed to extinction in a shorter timescale than the necessary for the system to find a suitable differentiated state.

In this work, we have focused on generic statistical properties of the model. Nevertheless, it has the potential to be applied to concrete situations where the tradeoff relationships are known or suspected. We developed the example of cyanobacteria. Interesting generalizations of this work would include more general situations, namely a fertility which is dependent φ_i-values. This would allow the scrutiny of systems that include exhibit tradeoffs between reproductive and somatic functions. This is relevant since it is believed that germ-soma tradeoff is important in the early stages of multicellularity. It would also allow us to consider the effect of different selection intensities at the somatic or reproductive level. Generalizations of the viability functions also allow us to include dependence on environmental factors, like the external nitrogen concentration in our example. This enables the study of different types of systems. One interesting application can be the study of the effect of seasonal external conditions in the differentiation. Returning to the example of cyanobacteria, does a periodical supply of external nitrogen favor or oppose specialization? What if the supply is random in time? The only reason for considering here all functions as equally important to the organism was of practical order. What happens in the case where different functions have different importances to the organism?

The main results presented in this chapter have been published in the article:

- Ref. [107]: The influence of the composition of tradeoffs on the generation of differentiated cells, André Amado, Paulo R. A. Campos, Journal of Statistical Mechanics (2017) DOI: 10.1088/1742-5468/aa71d8

4 A MECHANISTIC APPROACH TO TRADEOFFS AND COMPLEXITY

Highlights

A mechanistic model for multicellularity including tradeoffs and division of labor is introduced

Three different geometries for the aggregates are considered: linear, spherical and snowflake-like

The model is applied to the problem of the size-complexity rule and it is found that the validity of the rule depends on the geometry of the aggregates

The dynamics of group formation and evolution is governed at a microscopical level by a set of processes, which either increase or decrease group size. As such, it is useful to have a model description of group formation written directly in terms of these microscopic mechanisms that dictate group dynamics. Such a model can help us elucidate the phenomena driving the evolution of group size and complexity. Some work has been done in this domain. For example, a purely aggregative model without input generates in the long run one group only, comprising all the cells initially introduced in the system. A slightly more complex approach, accounting for aggregation and input processes, is analytically solvable and leads to a power law for the equilibrium distribution, with the number of groups of size n being proportinal to $n^{-3/2}$ [115], therefore lacking any characteristic size scale. In a 1995 paper, Gueron and Levin [116] modelled the dynamics of animal group formation in terms of density-dependent rates of fusion and fission. They focus on the mathematical properties of the model. Differently from the previously referred approaches, they restrict their analysis to continuous populations and do not include the processes of reproduction and death, employing fixed size populations.

A model aimed at emulating the evolution of cellular aggregates needs to give a fuller account of the processes involved. This way, our model seeks to provide an

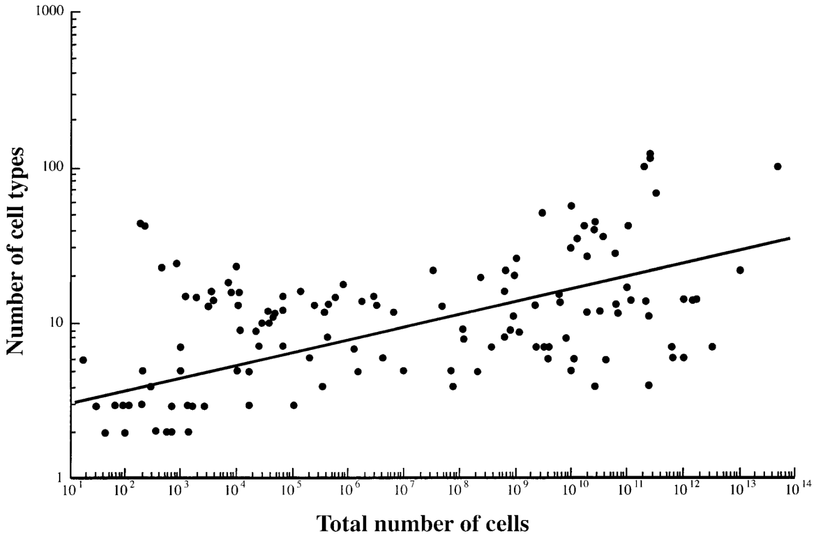

Figura 4.1: Number of cell types as a function of the number of cells in the organism for a vast range of species. Figure obtained from [118].

approach that can be used in practice to probe specific problems in the evolution of multicellular complexity, by incorporating four main processes: aggregation, dissociation, cell reproduction and cell death. Another important ingredient of the model is the presence of tradeoffs, since they are widely believed to play an essential role in the evolution of complexity [87]. We introduce the tradeoffs at the reproduction level, by considering that the cells should perform several tasks whose simultaneous execution brings high costs in terms of fitness.

Although the model can in principle be applied to a variety of problems, here we primarly intend to address the so called size-complexity rule. As described in chapter 1, the size-complexity rule consists in the observation that larger organisms tend to display higher complexity than smaller ones [117, 118]. There is no unique definition of complexity, but the number of cell types is frequently invoked as a proxy [119]. This rule seems to apply not only to multicellular organisms but also to other systems [118, 120], such as human societies. Nevertheless, this rule is not solidly established in theoretical grounds and there are reports of systems that appear to violate it.

Usually, cells separate completely upon cell reproduction but in situations where groups of cells are favored, cells may develop mechanisms that produce incomplete separation. These cells remain in their parent groups rather than dispersing. The evolution of this mechanism has been experimentally demonstrated in the laboratory, for example in [50]. To simulate this effect, we introduce a parameter σ dubbed stickiness that handles the probability of cells sticking together. In the limit

$\sigma \to 1$ cells always remain attached. Conversely, when $\sigma \to 0$ cells undergo perfect separation in every instance of cell reproduction.

We consider different aggregate geometries: a linear geometry, a spherical (or compact) geometry, and a snowflake geometry. These geometries are motivated by the ones observed in nature or laboratory experiments. In the linear geometry cells grow unidirectionally. In spherical geometry cells are considered to grow in such way that they form spherical cell aggregates, thus possessing a surface area proportional to $N^{2/3}$, where N is the number of cells in the aggregate. Finally, the snowflake geometry presents the highest complexity allowing for structures where each cell has up to z neighboring cells. The linear geometry can be seen as a particular case of the snowflake geometry with $z = 2$. The linear and spherical structures are simpler than the snowflake and their implementation does not need to keep track of the internal structure of the aggregate. On the contrary, the implementation of the snowflake structures needs to keep track of each individual member of the population, since there are many possible configurations that an aggregate of size ℓ can adopt. Examples of possible instances of these structures can be found in figure 4.2.

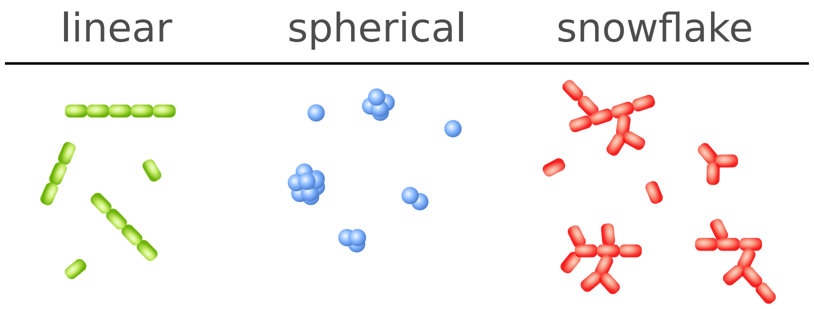

Figura 4.2: Example of structures with the geometries considered in the model.

In the following discussions, we borrow notation from chemical kinetics to represent the processes and their respective rates.

Notation summary

N_ℓ	Number of groups with size ℓ
N_T	Total number of cells
σ	Stickiness
k_+	Aggregation base rate

k_-	Dissociation base rate
k_R	Reproduction base rate
k_D	Death base rate
z	Maximum number of neighbors of a cell in snowflake structure
z_n	Number of neighbors of the cell n in snowflake structure

Aggregation

Groups of cells can merge in larger groups through the process of aggregation. This process can be described by the following expression

$$A_\ell + A_m \xrightarrow{k_+^{\ell,m}} A_{\ell+m} \tag{4.1}$$

where an aggregate of size ℓ and one of size m merge into a new aggregate of size $\ell + m$. For the linear process we adopt as aggregation rate $k_+^{\ell,m}$

$$k_+^{\ell,m} = k_+ \sigma^2 N_\ell N_m, \tag{4.2}$$

where k_+ is a constant of aggregation and N_ℓ the number of aggregates of size ℓ. As for the spherical model we assume that aggregation only occurs through the surface of the aggregate, so a factor of $\ell^{2/3}$ is included to reflect the surface area of a spherical aggregate. Also, in the spherical case we only consider aggregation of single cells and groups of variable sizes ℓ. Therefore we have

$$k_+^{\ell,1} = k_+ \sigma^2 N_\ell N_1 \ell^{2/3}. \tag{4.3}$$

Dissociation

Aggregates of cells can break into smaller groups. We consider the breakdown of a chain in case of linear structure or the removal of a cell from a spherical aggregate. This process can be described by

$$A_n \xrightarrow{k_-^n} A_\ell + A_m \tag{4.4}$$

where k_-^n stands for the rate of the process. In the linear structure this rate is given by

$$k_-^n = k_- N_n(n-1), \tag{4.5}$$

and in the spherical by

$$k_-^\ell = k_- N_\ell \ell^{2/3}. \tag{4.6}$$

Again k_- represents the base rate of dissociation. The linear structures have a factor $n-1$, representing the number of points where a linear structure can break, whereas in the spherical model a factor of $\ell^{2/3}$ is considered, proportional to the surface area of the agregate, through which cells can be lost.

Cell reproduction

The cells undergo cell reproduction with a rate k_R^ℓ, determined by the cells optimal reproduction rate R^{opt}. In the process of reproduction an aggregate of size l can produce an aggregate of size $l+1$, with probability σ^2, or produce a new free cell otherwise. This process can be represented by

$$A_\ell \xrightarrow{k_R^\ell \sigma^2} A_{\ell+1} \tag{4.7}$$

or

$$A_\ell \xrightarrow{k_R^\ell (1-\sigma^2)} A_\ell + A_1. \tag{4.8}$$

The rate k_R^ℓ is related to the number of cells in the aggregate ℓ as well as to the per capita reproduction rate R_ℓ^{opt}

$$k_R^\ell = k_R N_\ell R_\ell^{opt} \ell, \tag{4.9}$$

where k_R is a constant used to adjust the general reproduction rate and N_ℓ represents the total number of aggregates of size ℓ. The per capita reproduction rate functional form will be dealt with in Sec. 4.1 and Sec. 4.2.

In the snowflake structure, we consider that the cell is added as a neighbor of a cell that currently has less than the maximum number of neighbors z.

Cell death

Cells can die randomly with a uniform probability, proportional to the total number of cells in the system N_T. As such the death rate of a cell in an aggregate of size ℓ is simply given by

$$k_D^\ell = k_D N_\ell \ell N_T. \tag{4.10}$$

This corresponds to different processes depending on the geometry of the aggregate. When the m-th cell of a linear aggregate dies we have

$$A_\ell \xrightarrow{k_D^\ell} A_{\ell-m} + A_{m-1}. \tag{4.11}$$

When a cell in the spherical aggregate dies we simply have a reduction of the aggregate size by one

$$A_\ell \xrightarrow{k_D^\ell} A_{\ell-1}. \tag{4.12}$$

The case of snowflake structures is more delicate. In this situation when a central cell dies the aggregate can produce up to z daughter aggregates depending on the number of neighbors $z_n = 1, ..., z$ possessed by the dying cell. This process is illustrated in Fig. 4.3.

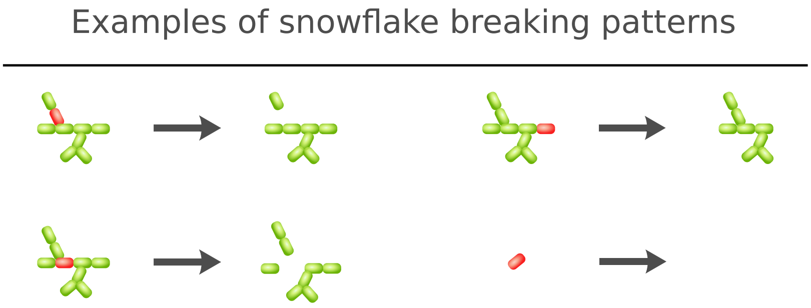

Figura 4.3: Example structures formed by breaking parent snowflake structures through cell death. The red cell is the one signaled to die.

4.1 Two-functions model

Reproduction rate

We can now include the effect of the tradeoffs in the cells' reproduction rate. We start by considering a single cell which needs to perform two essential incompatible functions, X and Y. To encode this in a mathematical language one can decompose the reproduction rate in two parts: a benefit B provided by the function and a cost C associated to it. As both functions are considered essential we choose to consider the benefit provided as the product of their expression levels. Thus, two functions with expression levels x and y provide a benefit $B(x, y)$

$$B(x, y) = x \cdot y. \tag{4.13}$$

As we want the total reproduction rate of the cell to be positive and finite, the cost should start slower but accelerate faster than the benefit, so that they intercept for some intermediate level of expression. A suitable function is a polynomial of degree higher than 2, since the benefit is a homogeneous function of degree 2. For simplicity we choose $C(x, y) = x^3 + y^3$. This function is well suited to describe a situation where no tradeoff is present, since the cost of performing function X is independent of the expression of function Y and vice-versa. To deal with this we can introduce a rapidly growing function that raises the costs of function X with the expression of Y. A simple example of such a function is

$$C(x, y) = x^3 e^{y^2} + y^3 e^{x^2}. \tag{4.14}$$

Other functions could have been chosen that reproduce the general qualitative behavior intended. Now clearly the cost increases significantly when both function are performed simultaneously, while keeping a much lower cost for the situation where the functions do not occur concurrently since $C(0, y) = y^3$ and $C(x, 0) = x^3$, which captures the tradeoff we were trying to encode in the function. The total reproduction rate of such an organism is therefore given by

$$R(x, y) = x \cdot y - c\left(x^3 e^{y^2} + y^3 e^{x^2}\right). \tag{4.15}$$

The introduction of a parameter c allows us to control the intensity of the cost/benefit relation.

This function can be easily generalized for n-cell aggregates where we consider that the function benefit is shared by all the cells in the aggregate. This way the per capita reproduction rate is

$$R(\boldsymbol{x}, \boldsymbol{y}) = \bar{x} \cdot \bar{y} - \frac{c}{n} \sum_{i=1}^{n} \left(x_i^3 e^{y_i^2} + y_i^3 e^{x_i^2} \right), \tag{4.16}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ are the average contributions of each cell to the aggregate. A similar model, for the case of single cells and two cell aggregates only, has been suggested by [109].

Let us now analyze the reproduction rate of an organism produced by such function. For one cell, the reproduction rate displays a single peak for an intermediate expression level of each function, which is expectable since both functions are essential (cf. Fig. 4.4 (a)). For two cells it is more difficult to visualize the resulting landscape, but we can take advantage of the symmetry of the function to understand that the optimum fitness should have the form $\boldsymbol{x} = (x, y)$ and $\boldsymbol{y} = (y, x)$. Thus we can restrict the search to a two dimentional piece of the parameter space. Doing this one can find that the optimal expression of the functions occurs when each cell specializes in a given function, instead of each of them performing both functions (cf. Fig. 4.4 (b)). This assumption of symmetry has been verified by a numerical study of the whole parameter space in the case of two cells.

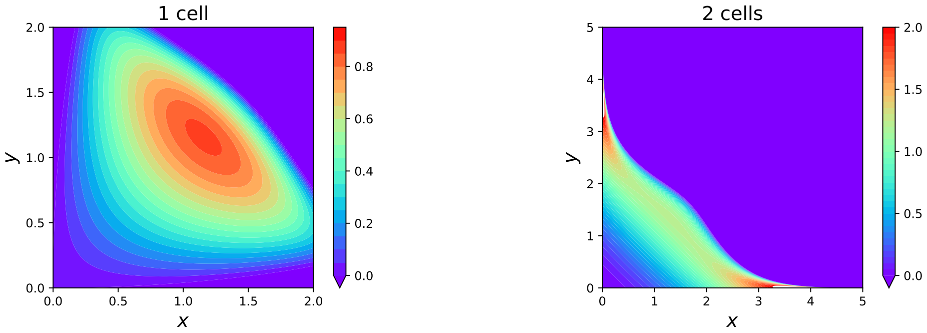

Figura 4.4: Left panel: reproduction rate for a single cell. Right panel: reproduction rate for a 2 cells aggregate. The cost c is $1/25$.

It is possible explore in more detail the effect of the tradeoff introducing a parameter a in the exponential of the cost function that allows to control the intensity of the tradeoff $C(x, y; a) = x^3 e^{ay^2} + y^3 e^{ax^2}$. This way it is possible to tune the strength of the tradeoff by varying a. Fig. 4.5 shows that when $a \to 0$ (no tradeoff) the values of x and y that produce the maximum fitness are equal. As the tradeoff is increased we achieve a critical value from which it is advantageous to the organism to start segregating the functions to different cells, leading to cell specialization. Mathematically, the maximum point corresponding to equal x and y becomes a saddle point and two new maxima arise in symmetrical positions. After total specialization the cells are not directly affected by the tradeoff anymore.On the other hand, if specialization is not allowed the reproduction rate keeps falling. This difference provides a strong push towards specialization, since any cell able to specialize will have a significant advantage in terms of reproduction rate. This allows the definition of a region of weak tradeoff, with $a \ll 1$, and a region of strong tradeoff, corresponding to $a \gtrsim 1$. As we are interested in studying the effect of strong tradeoffs in the evolution of cell aggregates we will consider $a = 1$ unless specified otherwise.

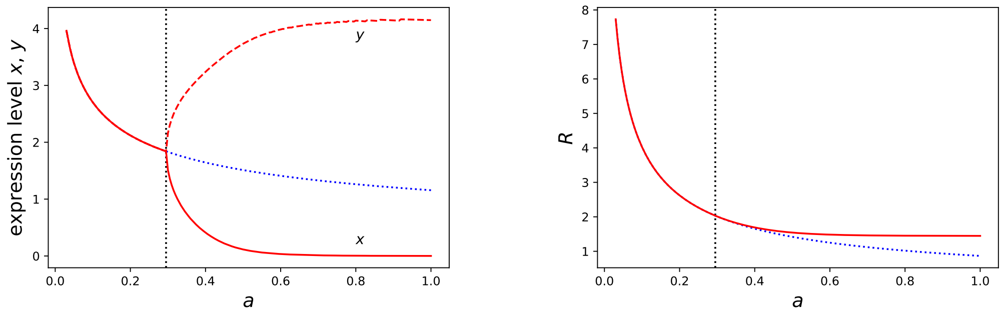

Figura 4.5: Left panel: value of the function expression x and y that maximize the reproduction rate as a function of the tradeoff parameter a. Right panel: reproduction rate as a function of the tradeoff parameter a. The red curves correspond to the free situation and the blue curves to a situation where x and y were made equal, preventing cell specialization. The vertical dotted line corresponds to the critical value of a for which differentiation starts to be favored ($a_c \approx 0.29$). The cost c is $1/25$.

The dependence of the optimized reproduction rate on the aggregate size is shown in Fig. 4.6 for cost parameter c equals 0.04, as an example.

4.2 Generalization to p-functions

As a next step in this study it is possible to generalize the model to deal with an arbitrary number of functions, lifting the restriction of two functions. This is important since it allows a clearer study of the relationship between size and complexity, here measured through the number of different functions an organism should perform.

The simplest way of extending the benefit to a situation with p-functions is to consider the product of the expression levels of all the p functions. This way

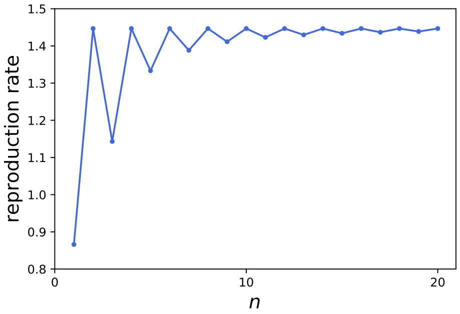

Figura 4.6: Reproduction rate as a function of the number of cells in the aggregate. The cost parameter c is 0.04.

the benefit of a group of n cells performing p functions would consist of $B_{(n,p)} = \prod_{\mu=1}^{p} \bar{x}_{(\mu)}$, where again $\bar{x}_{(\mu)} = \frac{1}{n}\sum_{i=1}^{n} x_{(\mu),i}$ is the average contribution of each of the n cells to the aggregate. Here $x_{(\mu)}$ represents the expression level of the function $X_{(\mu)}$, like x and y previously represented the functions X and Y. However, this function is not ideal since it is not obvious how to compare the result to the 2-celled situation or between any different number of functions. We adopt a modification of this function that allows for a clearer comparison between different number of functions. The new benefit is defined to be

$$B_{(n,p)} = \left(\prod_{\mu=1}^{p} \bar{x}_{\mu}\right)^{2/p}. \tag{4.17}$$

This definition guarantees that the benefit function is still a second order homogeneous function regardless the number of functions performed by the organism.

The following step is to find a suitable function to represent the cost. We consider that the tradeoffs occur between pairs of functions and thereby a simple generalization of the cost is

$$C_{(n,p)} = \frac{c_p}{n}\sum_{i=1}^{n}\sum_{\mu=1}^{p} x_{(\mu),i}^{3} \exp\left(\sum_{\nu\neq\mu} x_{(\nu),i}^{2}\right). \tag{4.18}$$

With this function all the different functions performed in the same cell contribute to the tradeoff. We have kept an index p in the constant c because we shall determine later how the costs of different functions relate in such a way that the maximum performance remains fixed, independently of the number of functions p.

This benefit and cost functions produce a reproduction rate that reduces to the previous case of 2 functions when p is set to 2. The reproduction rate is thus

$$R_{(n,p)} = B_{(n,p)} - C_{(n,p)} = \left(\prod_{\mu=1}^{p} \bar{x}_\mu\right)^{2/p} - \frac{c_p}{n}\sum_{i=1}^{n}\sum_{\mu=1}^{p} x^3_{(\mu),i} \exp\left(\sum_{\nu\neq\mu} x^2_{(\nu),i}\right). \tag{4.19}$$

To allow a comparison between different number of functions we consider that the maximum fitness $R^{max}_{(mp,p)}$ (when there is total specialization) is equal. For that purpose, the constant c_p should be adjusted. Since we are working in a region of strong tradeoff, the maximum reproduction rate is achieved for total specialization. Moreover, the maximum is achieved when the number of cells n is a multiple of the number of functions p. In this case we have

$$R^{opt}_{(mp,p)} = \left[\left(\frac{m\chi}{mp}\right)^p\right]^{2/p} - \frac{c_p}{mp}\left(mp\chi^3\right) = \left(\frac{\chi}{p}\right)^2 - c_p\chi^3 \tag{4.20}$$

where $n = mp$ is the number of cells, taken to be an integer multiple of p and χ is the optimal value for the task expression. Notice that, due to total specialization of the cells, the exponential factors disappear. The value χ which optimizes this expression is then

$$\left.\frac{\mathrm{d}R^{opt}_{(mp,p)}}{\mathrm{d}\chi}\right|_{\chi=\chi_{opt}} = 0 \quad \Rightarrow \quad \frac{2\chi}{p^2} - 3c_p\chi^2 = 0 \quad \Rightarrow \quad \chi = 0 \vee \chi = \frac{2}{3}\frac{1}{p^2\,c_p} \tag{4.21}$$

The value of $\chi = 0$ is obviously a minimum of the reproduction rate and so $\chi_{opt} = \frac{2}{3}\frac{1}{p\,c_p}$ is the only interesting value to consider. Replacing χ_{opt} in the expression for the reproduction rate we are left with

$$R^{opt}_{(mp,p)} = \frac{4}{9}\frac{1}{p^6\,c_p^2} - \frac{c_p}{p^6}\frac{8}{27}\frac{1}{c_p^3} = \frac{4}{27}\frac{1}{p^6 c_p^2}. \tag{4.22}$$

Now, guaranteeing that this rate remains equal for different p implies that

$$R^{opt}_{(mp,p)} = R^{max}_{(m'p',p')} \Rightarrow \frac{4}{27}\frac{1}{p^6 c_p^2} = \frac{4}{27}\frac{1}{p'^6 c_{p'}^2} \Rightarrow c_p = \frac{p'^3}{p^3}c_{p'} \tag{4.23}$$

We choose to take $p = 2$ as a reference ($c_2 \equiv c$) so we have $c_p = \frac{8}{p^3}c_2 = \frac{8}{p^3}c$. This normalized reproduction rate is finally

$$R_{(n,p)}^{Normalized} = \left(\prod_{\mu=1}^{p} \bar{x}_\mu\right)^{2/p} - \frac{8c}{p^3 n} \sum_{i=1}^{n} \sum_{\mu=1}^{p} x_{(\mu),i}^3 \exp\left(\sum_{\nu\neq\mu} x_{(\nu),i}^2\right). \tag{4.24}$$

Now that we have introduced the general expression, let us analyze the fitness optimum values for different number of cells. When the number of cells is a multiple of the number of functions it is possible to distribute the workload perfectly among the cells and the fitness can achieve its maximum value. In this case we can write $n = mp$ and the reproduction rate corresponds to the previously calculated optimum one

$$R_{(mp,p)}^{*} = \frac{4}{27p^6c_p^2} = \frac{1}{432c^2} \tag{4.25}$$

where the asterisk is used to specify that it refers to the optimal value of the reproduction rate for the given number of cells and functions. Notice that the value becomes independent of the number of cells of functions in accordance with our previous normalization choice.

4.2.1 General $n \geq p$

Now let us turn to the situation where the number of cells is not a multiple of p anymore, rendering a perfect division of labour impossible. We will restrict for the moment to the situation where $n > p$. In such a case, we can write $n = mp + k$ where m is a non-zero integer and k an integer in the range $0 < k < p$. An exaustive analysis of the structure of the expression of the task showed us that, in the parameter region of interest, the best configuration has the following structure

$$\begin{cases} (m+1)\text{ cells} & \rightarrow \text{perform activity } X_{(1)} \text{ at a level } x\,, \\ (m+1)\text{ cells} & \rightarrow \text{ perform activity } X_{(2)} \text{ at a level } x\,, \\ \quad\vdots & \qquad\vdots \\ (m+1)\text{ cells} & \rightarrow \text{ perform activity } X_{(k)} \text{ at a level } x\,, \\ m\text{ cells} & \rightarrow \text{ perform activity } X_{(k+1)} \text{ at a level } y\,, \\ m\text{ cells} & \rightarrow \text{ perform activity } X_{(k+2)} \text{ at a level } y\,, \\ \quad\vdots & \qquad\vdots \\ m\text{ cells} & \rightarrow \text{ perform activity } X_{(p)} \text{ at a level } y\,, \end{cases} \tag{4.26}$$

i.e., k functions are performed by $m+1$ cells at expression level x, while $p-k$ functions are performed by m cells at level y. This structure results in a reproduction rate of

$$\begin{aligned} R^{*}_{(mp+k,p)} = & \left[\left(\frac{(m+1)\,x}{mp+k} \right)^{k} \left(\frac{m\,y}{mp+k} \right)^{(p-k)} \right]^{2/p} \\ & - \frac{8\,c}{p^{3}\,n} \left[k(m+1)\,x^{3} + (p-k)m\,y^{3} \right] \end{aligned} \tag{4.27}$$

for which remains to find the optimal values of x and y. In the optimum, we have

$$\vec{\nabla} R(x,y)\Big|_{x=\chi,y=\gamma} = 0 \quad \Rightarrow \quad \vec{\nabla} B(x,y)\Big|_{x=\chi,y=\gamma} - \vec{\nabla} C(x,y)\Big|_{x=\chi,y=\gamma} = 0 \tag{4.28}$$

So we should compute the gradients of B and C

$$\frac{\partial B}{\partial x} = \frac{2k}{p} \left[\frac{(m+1)^{k}(my)^{p-k}}{n^{p}} \right]^{2/p} x^{2k/p-1} = \frac{2k}{p}\frac{B}{x}, \tag{4.29}$$

$$\frac{\partial B}{\partial y} = \frac{2(p-k)}{p} \left[\frac{((m+1)x)^{k} m^{p-k}}{n^{p}} \right]^{2/p} y^{2(p-k)/p-1} = \frac{2(p-k)}{p}\frac{B}{y} \tag{4.30}$$

and

$$\frac{\partial C}{\partial x} = \frac{8c}{np^{3}}\, 3k(m+1)x^{2}, \tag{4.31}$$

$$\frac{\partial C}{\partial y} = \frac{8c}{np^{3}}\, 3(p-k)my^{2}. \tag{4.32}$$

Therefore,

$$\begin{cases} \frac{2k}{n^{2}p} \left[((m+1)\chi)^{k}\,(m\gamma)^{p-k} \right]^{2/p} & = \frac{8c}{np^{3}}\, 3k(m+1)\chi^{3}, \\ \frac{2(p-k)}{n^{2}p} \left[((m+1)\chi)^{k}\,(m\gamma)^{p-k} \right]^{2/p} & = \frac{8c}{np^{3}}\, 3(p-k)m\gamma^{3} \end{cases} \tag{4.33}$$

where we multiplied both sides of the top equation by χ and both sides of bottom equation by γ. Simplifying leads to

$$\begin{cases} \frac{1}{n} \left[((m+1)\chi)^{k}(m\gamma)^{p-k} \right]^{2/p} & = \frac{12c}{p^{2}}\,(m+1)\chi^{3}, \\ \frac{1}{n} \left[((m+1)\chi)^{k}(m\gamma)^{p-k} \right]^{2/p} & = \frac{12c}{p^{2}}\,m\gamma^{3}. \end{cases} \tag{4.34}$$

As the left hand side of both equation is equal we can find

$$\frac{12c}{p^2}(m+1)\chi^3 = \frac{12c}{p^2}m\gamma^3 \Rightarrow \boxed{(m+1)\chi^3 = m\gamma^3}. \tag{4.35}$$

It is now possible to replace this value of χ in the first equation to obtaining

$$\left[(m+1)^k\left(\frac{m}{m+1}\right)^{k/3}\gamma^k m^{p-k}\gamma^{p-k}\right]^{2/p} = \frac{12nc}{p^2}m\gamma^3 \tag{4.36}$$

$$\tag{4.37}$$

which leads to

$$\gamma = \frac{m\,p^2}{12nc}\left(\frac{m+1}{m}\right)^{\frac{4k}{3p}} \tag{4.38}$$

and, consequently,

$$\chi = \frac{m\,p^2}{12nc}\left(\frac{m+1}{m}\right)^{\frac{4k-p}{3p}}. \tag{4.39}$$

Knowing χ and γ one can easily find the values of the benefit and cost that optimize the growing rate by direct substitution

$$\begin{aligned} B(\chi,\gamma) &= \frac{1}{n^2}\left[((m+1)\chi)^k\,(m\gamma)^{p-k}\right]^{2/p} \\ &= \frac{1}{n^2}\left[(m+1)^k m^{p-k}\chi^k\left(\frac{m+1}{m}\right)^{\frac{p-k}{3}}\chi^{p-k}\right]^{2/p} \\ &= \frac{(m+1)^{\frac{2(p+2k)}{3p}}m^{\frac{4(p-k)}{3p}}}{n^2}\chi^2 \\ &= \frac{(m+1)^{\frac{2(p+2k)}{3p}}m^{\frac{4(p-k)}{3p}}}{n^2}\left[\frac{mp^2}{12nc}\left(\frac{m+1}{m}\right)^{\frac{4k-p}{3p}}\right]^2 \\ &= \frac{p^4}{144\,n^4c^2}(m+1)^{4k/p}m^{4(p-k)/p}. \end{aligned} \tag{4.40}$$

and

$$\begin{aligned}C(\chi,\gamma) &= \frac{8c}{np^3}\left[k(m+1)\chi^3+(p-k)m\gamma^3\right]\\ &= \frac{8c}{np^3}\left[k(m+1)\chi^3+(p-k)(m+1)\chi^3\right] = \frac{8c}{np^3}(m+1)\chi^3\\ &= \frac{8c}{np^3}(m+1)\frac{p^6m^3}{12^3c^3n^3}\left(\frac{m+1}{m}\right)^{(4k-p)/p}\\ &= \frac{2\,p^4}{432c^2n^4}(m+1)^{4k/p}m^{4(p-k)/p} = \frac{2}{3}B(\chi,\gamma). \end{aligned} \tag{4.41}$$

Therefore the total reproduction rate is

$$R^*_{(n=mp+k,p)}(\chi,\gamma) = B(\chi,\gamma)-C(\chi,\gamma) = \frac{1}{3}B(\chi,\gamma) = \frac{p^4(m+1)^{4k/p}m^{4(p-k)/p}}{432c^2n^4} \tag{4.42}$$

At the optimum, the cost is two thirds of the benefit. This apparent coincidence allows us to guarantee that the structure chosen in (4.26), indeed optimizes the reproduction rate, with the assumption of a total division of labor. This result is detailed below.

4.2.2 Relation between the benefit and the cost at the optimum

There is an interesting general result that can be shown at the optimum of the reproduction rate. The maximum of reproduction rate happens when

$$\vec{\nabla}R(\vec{x}) = 0 \quad \Rightarrow \quad \vec{\nabla}B(\vec{x}) - \vec{\nabla}C(\vec{x}) = 0 \quad \Rightarrow \quad \vec{\nabla}B(\vec{x}) = \vec{\nabla}C(\vec{x}). \tag{4.43}$$

The benefit function considered here is a homogeneous function of degree 2, which will allow us to take advantage of the Euler's Homogeneous Function Theorem, which guarantees that

$$\vec{x}\cdot\vec{\nabla}f(\vec{x}) = nf(\vec{x}) \tag{4.44}$$

given that $f(\vec{x})$ is a homogeneous function of degree n. This way, we can compute the inner product of equation (4.43) with $\vec{x}$, producing

$$\begin{aligned} \vec{x} \cdot \vec{\nabla} B(\vec{x}) &= \vec{x} \cdot \vec{\nabla} C(\vec{x}) \\ 2B(\vec{x}) &= \vec{x} \cdot \vec{\nabla} C(\vec{x}) \\ B(\vec{x}) &= \frac{1}{2} \vec{x} \cdot \vec{\nabla} C(\vec{x}) \end{aligned} \tag{4.45}$$

This result is analogous to the Virial Theorem in physics! The Virial Theorem for one particle in an arbitrary differentiable potential states that [121]

$$\langle K \rangle = -\frac{1}{2} \langle \vec{r} \cdot \vec{F} \rangle = \frac{1}{2} \langle \vec{r} \cdot \vec{\nabla} V \rangle \tag{4.46}$$

which is exactly what we find in our system. In this analogy, the benefit function plays the role of kinetic energy and the cost function the role of a potential energy. This result stems from the fact that kinetic energy is a homogeneous function of degree 2.

Now, in the special case of total division of labor the cost function becomes a homogeneous function of degree 3 since the exponentials disappear. As a consequence, we get $\vec{x} \cdot \vec{\nabla} C(\vec{x}) = 3C(\vec{x})$ and, therefore,

$$C(\vec{x}) = \frac{2}{3} B(\vec{x}). \tag{4.47}$$

This is the relation we found previously between cost and benefit, which further confirms that our *ansatz* captures the maximum of the function, under the assumption of total division of labor. Although the division of labor is probably not perfect, it is a quite good approximation in the regime of strong tradeoffs, that we are working on, as shown in Fig. 4.5 and by the exploration done in the parameter space.

4.2.3 $n < p$

Finally, it remains to analyze the case when the number of cells is smaller than the number of functions to be performed. In this situation it is not possible to achieve a total specialization of the cells since some tasks will be missing and the benefit function will vanish. We have obtained that, except at very low values of the cost parameter c, the configuration that leads to optimized conditions corresponds to the one in which all cells are generalists thus rendering all activities at the same level γ. This way the reproduction rate becomes

$$R^*_{(n,p)} = \left[\left(\frac{n\gamma}{n}\right)^p\right]^{2/p} - \frac{8c}{p^3 n}\left(np\gamma^3 e^{(p-1)\gamma^2}\right) = \gamma^2 - \frac{8c}{p^2}\gamma^3 e^{(p-1)\gamma^2}. \tag{4.48}$$

Notice that in this case the reproduction rate is independent of the number of cells in the aggregate. When maximizing this function we face a transcendental equation, that cannot be solved analytically. Therefore, we rely on a numerical solution in this case. Fig. 4.7 shows the optimal reproduction rate for several values of p. We can see that the reproduction rate monotonically decreases with p, which is expectable since there is no division of labor and as such the tradeoffs become stronger as we increase the number of tasks. Also expected is the decrease of the reproduction rate as the cost parameter c is increased.

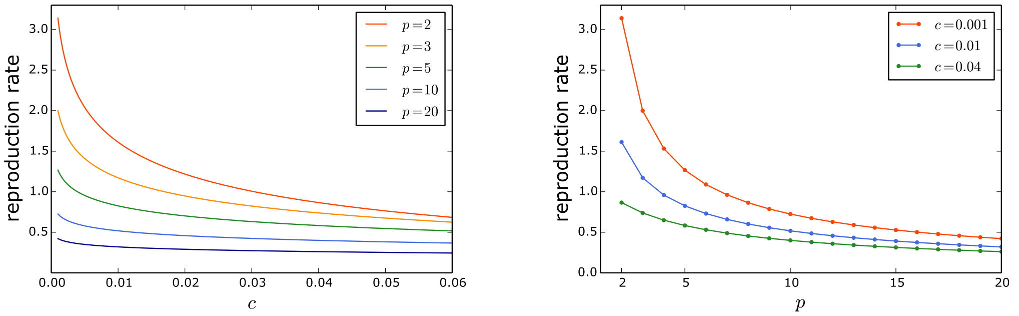

Figura 4.7: Left panel: Optimal reproduction rate for $n < p$ as a function of the cost parameter c, shown for 5 values of p. Right panel: Optimal reproduction rate for $n < p$ as a function of the number of tasks p, shown for three values of c.

4.2.4 General n

With these results we can finally put together the reproduction rate as a function of n for different number of tasks. Fig. 4.8 shows the reproduction rate used in the simulations as a function of n for different sets of parameters. In the right panel of Fig. 4.8 we can see that, as the cost is reduced, the reproduction rate for $n \geq p$ grows fast (proportionally to c^{-2}), enhancing the difference between specialized and generalist cells, which grow more slowly.

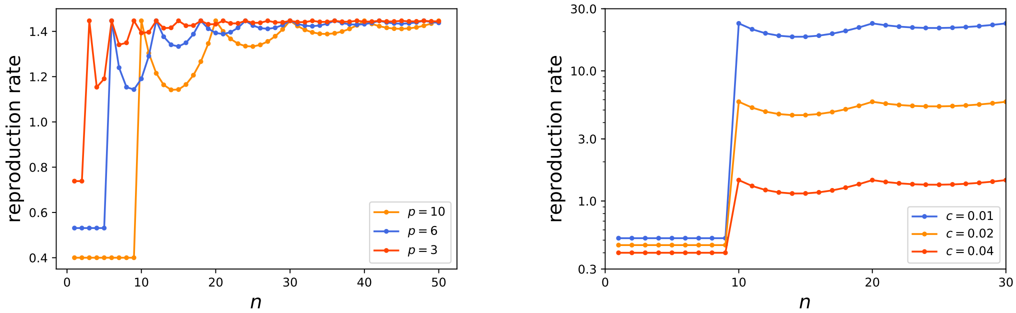

Figura 4.8: Left panel: Optimal reproduction rate as a function of the aggregate size n, shown for three values of p and $c = 0.04$. Right panel: Optimal reproduction rate as a function of the aggregate size n, shown for three values of c for $p = 10$.

4.3 Simulation results

Now that we have explored in detail the properties the reproduction rate we can proceed to the simulation of the model, including all four processes. Let us start with the two-functions model and later on move to the generalization to arbitrary p.

The simulation is initiated with a population of $10\,000$ single cells. This condition does not correspond to the equilibrium of the system. As such, the system evolves until an equilibrium is reached. Fig. 4.9 shows the evolutionary trajectory of average and maximum group sizes during the initial 10^6 iterates. One can see that, for the set of parameters depicted in the graph, equilibrium is reached after a transient period of approximately 10^5 iterates. The group sizes rapidly evolve from the initial size of 1 to relatively large groups, as cells group together, reproduce and die.

Once equilibrium is reached the population is characterized by a stationary distribution of group sizes. Fig. 4.10 displays the probability distribution of finding a group of a certain size at any moment after the equilibrium has been reached. Notice that the scale is logarithmic so the figures show a wide range of probabilities and aggregate sizes. The behavior of both models shares several common features. The increase in stickiness σ leads to the formation of larger aggregates, as expected. For large stickiness, the probability distribution of group sizes is nearly flat over a broad range of size and then abruptly drops when the group size passes a certain threshold. This cutoff size increases with the stickiness. From the bottom graphs of Fig. 4.10 we can tell that this drop is exponential and that the onset of the exponential decay in probability happens sooner in the linear model than in the spherical

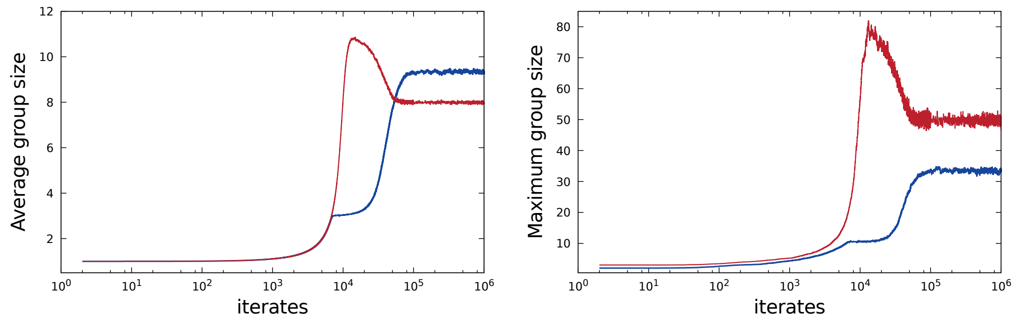

Figura 4.9: Evolutionary trajectories for mean group size (left panel) and largest group size (right panel) for the $p = 2$ model. The red lines represent the results for the linear model whereas the blue lines refer to the spherical model. The parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$, $k_D = 10^{-5}$, $\sigma = 0.5$ and $c = 0.04$. The results correspond to an average over 100 independent runs. Figure adapted from [122].

one. The dashed line in the top graphs represents the expected result for a model of pure aggregation with input. A simple model with aggregation and input only has been solved analytically, yielding a distribution proportional to $n^{-3/2}$ [115]. Our result is clearly distinct from the result of aggregation with input. Therefore it is clear that the remaining mechanisms play an essential role. Whereas the simple aggregation with input yields a power-law tail, our model produces an exponential one. This is important since some properties of power-laws, such as average size, are not well-defined. This does not allow us to define any characteristic sizes of the aggregates created. On the contrary, our model displays an exponential tail, which allow us to probe average aggregate sizes and average maximum aggregate sizes.

All these together evince that the increase observed in group size is not solely determined by the process of aggregation. Indeed, even the process of input, appearance of new cells, is ruled by the dynamics of the system itself. In this context, there is an interesting feedback between natural selection and growth dynamics of aggregates, as their size strongly influences the reproductive rate of cells.

For the sake of completeness, in Fig. 4.11 we present the distribution of average and maximum size at the equilibrium. The average sizes at equilibrium approximately follow a gaussian distribution, symmetric around the mean value, while the maximum size display a lognormal behavior, with a longer tail to large sizes.

In order to study the role of division of labor itself we now compare the original model with a modified version in which cells do not specialize. In this modified

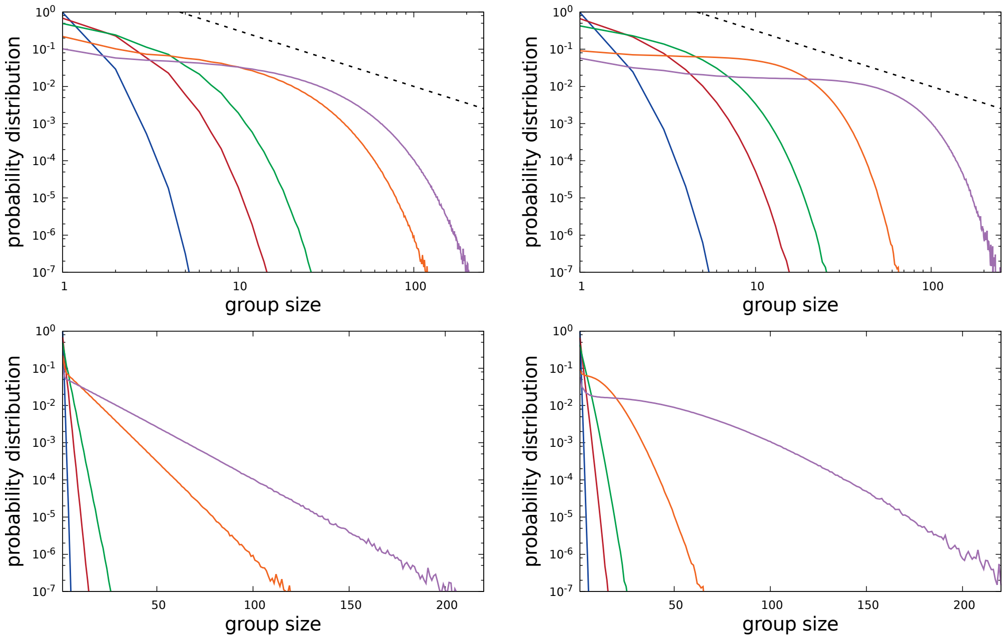

Figura 4.10: Probability distribution of group sizes for the linear (left) and spherical (right) models. The top figures are presented in logarithmic scale in both probability and size axes, while only the probability axis is in logarithmic scale for the bottom figures. The values for stickiness are from left to right 0.01, 0.05, 0.1, 0.5, 0.9. The remaining parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$, $k_D = 10^{-5}$, and $c = 0.04$. The results correspond to an average over 100 independent runs. The straight dashed line serves as an eye-guide describing a power-law with exponent $-3/2$, which is expected to describe the tail of the distribution in a system with only the mechanism of aggregation with input. Figure adapted from [122].

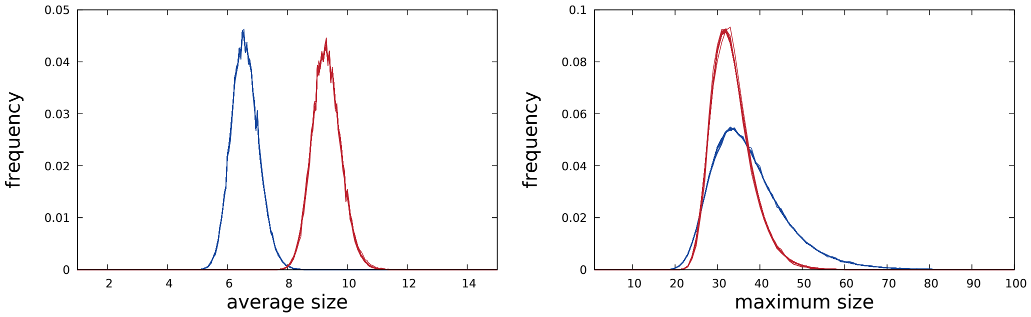

Figura 4.11: Distribution of average size (left) and maximum size (right) at the equilibrium. The red lines represent the linear model whereas the blue lines represent the spherical one. The parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$, $k_D = 10^{-5}$, $\sigma = 0.5$ and $c = 0.04$. The results correspond to an average over 100 independent runs.

model all cells are regarded as generalists, thus performing both tasks at a constant rate. More explicitly, we consider the following configuration

$$x_1 = y_1 = \cdots = x_n = y_n = \lambda \quad \Rightarrow \quad R_n(\mathbf{x}, \mathbf{y}) = \lambda^2 - 2c\,\lambda^3 e^{\lambda^2}. \tag{4.49}$$

We take the value of λ that maximize the reproduction rate in this generalist configuration. Since this equation is a transcendental equation we rely on a numerical solution. As one can conclude from equation 4.49, in this generalist form the reproduction rate becomes independent of the number of cells since they cannot benefit anymore from the division of labor. The results obtained are summarized in Fig. 4.12. In the plot we make $k_+ = k_- = k$. We can see that the division of labor leads to a different group structure, which is more obvious in the middle graphs that show the ratio between the probability values for a given group size. Division of labor always leads to greater group sizes, specially for low values of k. As we increase k, the effect of reproduction becomes more diluted by the processes of aggregation and dissociation and the relevance of division of labor is reduced. Bottom panels in Fig. 4.49 display a zoom of the middle panels for low group size. It is interesting to notice that the ratio between probability values with and without division of labor presents an up-and-down pattern with period two. This behavior is not due to noise. Instead, it reflects the fact that the reproduction rate for the specialized cells model oscillates (check Fig. 4.6 for details) while the generalist's reproduction rate is constant.

Fig. 4.13 reveals the dependence of the average and largest group sizes on the stickiness σ and the cost parameter c. We clearly see that a higher cost of performing the task yields smaller sized groups. Intuitively, higher stickiness favors larger groups. For low values of stickiness, the cost parameter c, plays a minor role and the average group size becomes roughly constant. This is expected to be exact for $\sigma = 0$ since in this limit all groups should be composed of a single cell. Although the average group sizes are not quite distinct in the two models, in the linear model the largest groups can be comprised of a relatively higher number of cells, like one hundred or more cells, as achieved for low cost c and high stickiness σ. One important remark is that for low cost c, the formation of groups of a given size requires a minimal level of stickiness σ that is larger for the spherical model.

The dependence of both average and largest group size on the dissociation parameter k_- and on the aggregation parameter k_+ are displayed in Fig. 4.14. In the linear model (left panel), the variables k_+ and k_- have opposing effect on each quantity. While group size grows as the aggregation rate k_+ increases, it shrinks

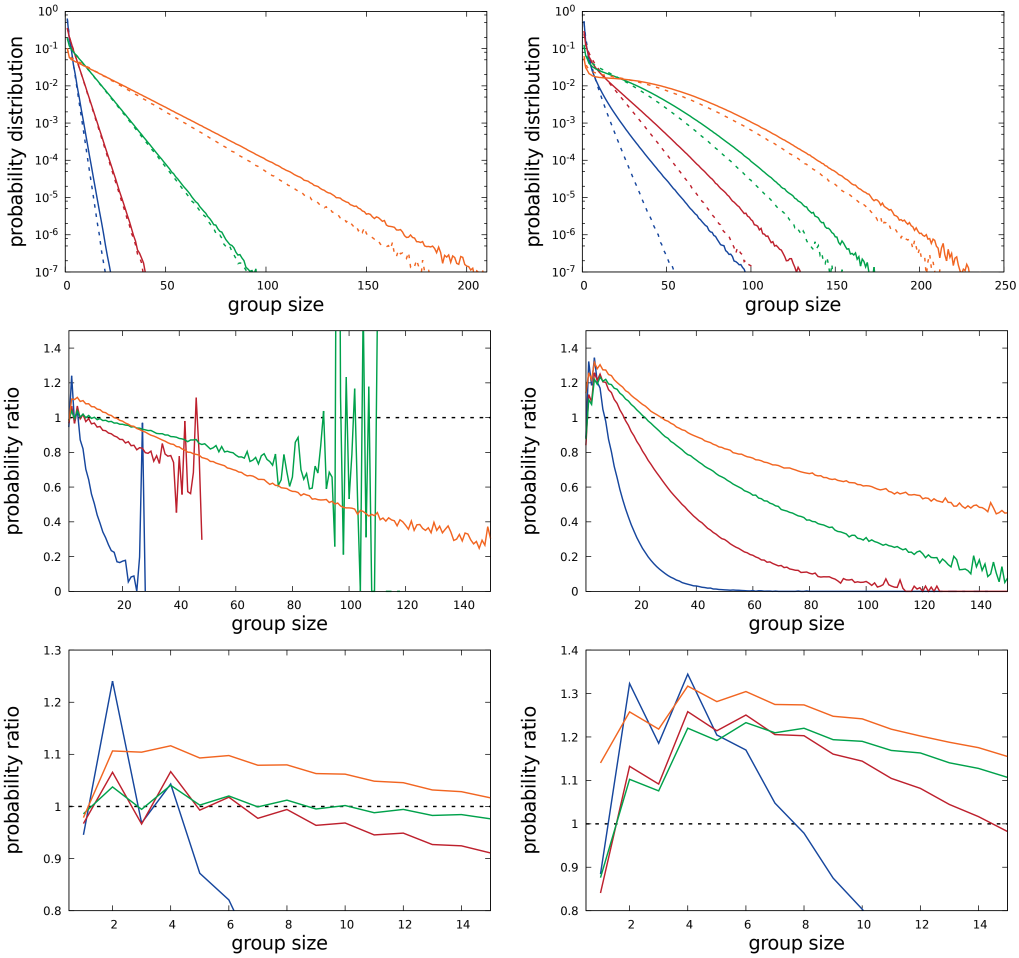

Figura 4.12: Top panels: Probability distribution of group sizes for the linear (left) and spherical (right) models. The full lines correspond to results of the original model (specialist cells), whereas the dashed-lines correspond to the case in which all cells perform both tasks at the same level (generalist cells). Middle and bottom panels: Ratio between the probabilities of the generalist cells and specialist cells for the linear (left) and spherical (right) models. The values for k_+ and k_- are from left to right $k_+ = k_- = 0$, 10^{-4}, 10^{-3} and 10^{-2}. The remaining parameter values are $\sigma = 0.9$, $k_R = 0.01$, $k_D = 10^{-5}$, and $c = 0.04$. The results correspond to an average over 250 independent runs.

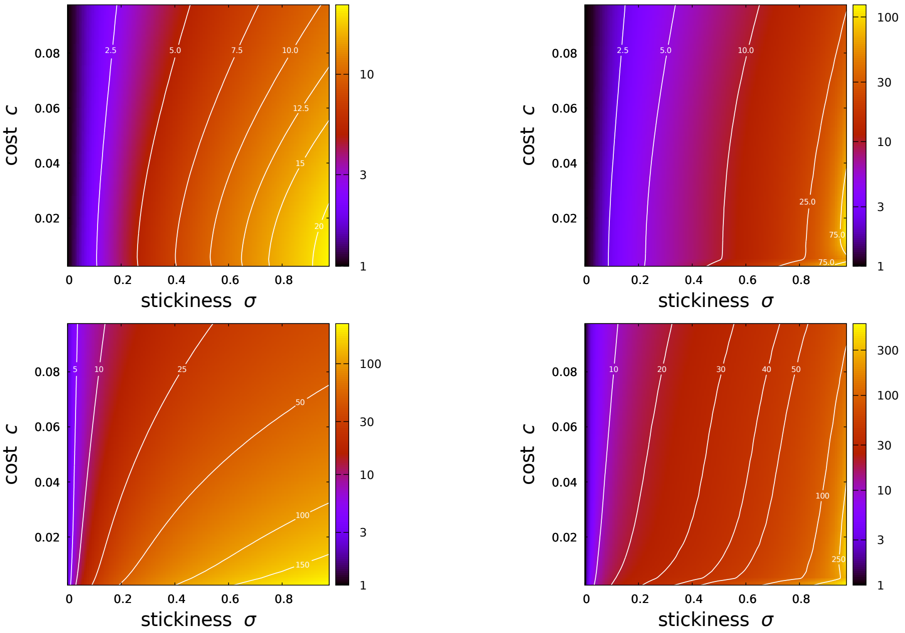

Figura 4.13: Average group size (upper panels) and maximum group size (lower panels) for the linear (left) and spherical (right) models as a function of σ and c. The remaining parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$ and $k_D = 10^{-5}$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 16 independent runs. Figure adapted from [122].

with the increase of the dissociation rate k_-. For fixed k_+, the increase of k_- always results in smaller groups, as we easily notice from the isoclines. For small values of k_+ a change in the rate of dissociation k_- has a minor effect since the group size is anyway quite small. On the other hand, the spherical model exhibits a more counterintuitive behavior. While for small values of the aggregation rate k_+, the increase in the dissociation rate leads to the formation of smaller groups, for intermediate and large values of k_+ the average group size peaks at intermediate values of k_-. This feature is highlighted when analyzing the maximum group size for the spherical model, which clearly reveals that the increase of k_- at large k_+ can produce larger groups. This effect is probably due to the fact that increasing k_- leads to a higher number of free cells that are available to be incorporated in the growth of other groups.

It remains to analyze the behavior of the number of groups. This study is presented in Figs. 4.15 and 4.16. Let us first focus on the dependence of the number of groups with stickiness σ and cost parameter c, presented in Fig. 4.15. In general, one can observe that larger stickiness at fixed cost leads to smaller number of groups. On the other hand, an increased cost c implies a smaller population of ag-

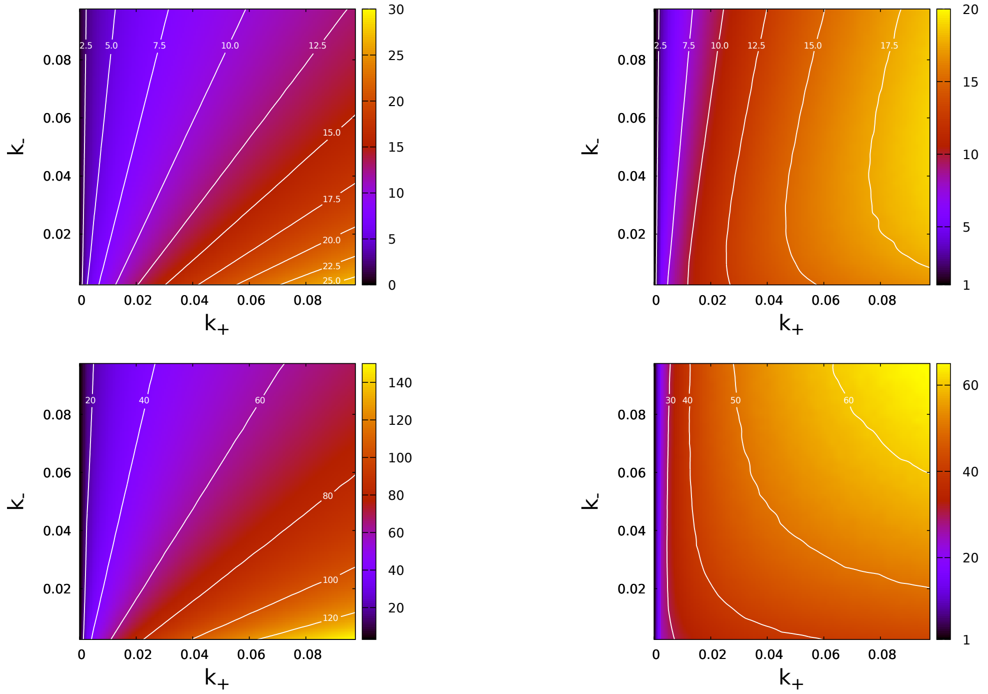

Figura 4.14: Average group size (upper panels) and maximum group size (lower panels) for the linear (left) and spherical (right) models as a function of k_+ and k_-. The remaining parameter values are $k_R = 0.01$, $k_D = 10^{-5}$, $\sigma = 0.5$ and $c = 0.04$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 16 independent runs. Figure adapted from [122].

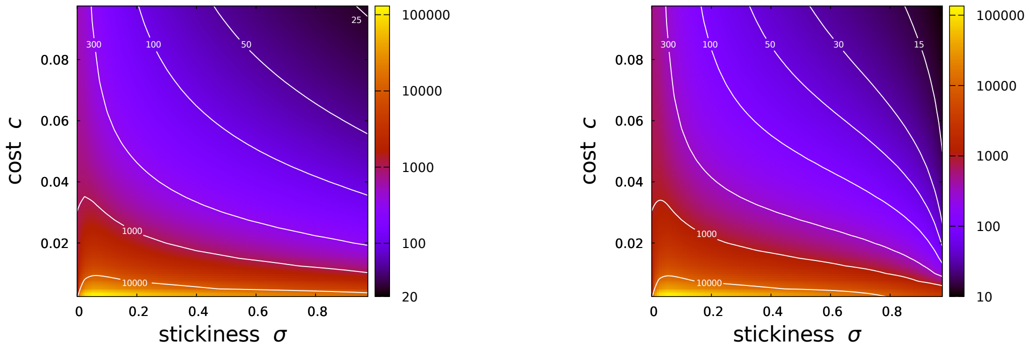

Figura 4.15: Number of groups for the linear (left) and spherical (right) models as a function of σ and c. The remaining parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$ and $k_D = 10^{-5}$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 16 independent runs. Figure adapted from [122].

gregates at stationarity. A smaller number of groups with increased tradeoff costs is in agreement with previous studies [96, 107]. Except for small cost c, the number of aggregates peaks at small stickiness σ. The reader is referred to Fig. 4.16, for information regarding the dependence of the number of groups on the dissociation k_- and aggregation k_+ rates. Linear groups present a relatively straightforward relationship, with the number of groups rising with k_- and decreasing with k_+. Once again, the spherical model displays a subtler scenario in comparison to the linear model. At intermediate and large k_+ the number of aggregates attains its minimum for intermediate values of k_-. This is complementary to the outcome presented in Fig. 4.14, as this region reflects the emergence of larger groups. For both models, the highest number of groups is found for smaller values of k_+. It is worth to stress that in the graph σ is set to 0.5, leading to a probability of cells staying together after reproduction of only 0.25. This way, under this conditions, aggregation is an important mechanism to drive the increase of group size. Therefore, when we make the aggregation rate k_+ very small, the groups will become smaller, leading to a fragmentation of the population into numerous small groups.

4.3.1 Arbitrary tradeoffs

Having analyzed the case of two tradeoffs we turn our attention to a more general case, consisting of an arbitrary number of tasks. This situation is quite interesting since it provides us with a tool to probe the effect of complexity, measured through the number tasks to perform, on the size of the cellular aggregates. This model opens a new powerful way for studying the size-complexity rule.

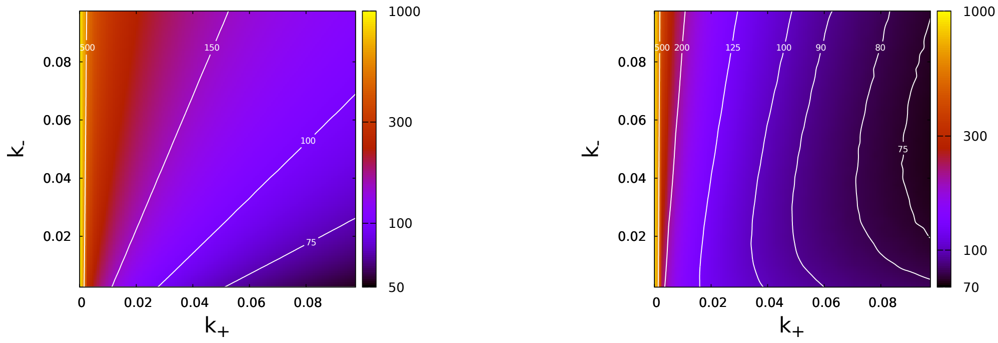

Figura 4.16: Number of groups for the linear (left) and spherical (right) models as a function of k_+ and k_-. The remaining parameter values are $k_R = 0.01$ and $k_D = 10^{-5}$, $\sigma = 0.5$ and $c = 0.04$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 16 independent runs. Figure adapted from [122].

Additionally, here we will also study two distinct life cycles and their roles on the evolutionary dynamics: the aggregative development and the clonal development. The aggregative life cycle contains all 4 kinds of processes we introduced. On the other hand, the clonal development can be approached by making the aggregation and dissociation rates equal zero, $k_+ = k_- = 0$, and restrict the study to high values of stickiness σ.

4.3.2 Aggregative development

Aggregative development is a mode of development where cells from different sources group together. This mode of development is also known as "coming together" in the literature [123]. In this stage we consider nonzero rates of aggregation and dissociation, k_+ and k_-. These processes play an important role in the aggregative mode of development but they are not exclusive, as reproduction and death play equally an important role in the dynamics. The outcomes of this model clearly differ from simple models of aggregation and dissociation [115, 122]. Indeed the processes of "staying together" are highly relevant at high stickiness and the simultaneous occurrence of the processes of "coming together" and "staying together" is allowed [123, 124].

As previously stated, we are interested in studying the size-complexity rule in the context of the model. Therefore we should study the impact of the complexity of the aggregates on their typical sizes. This is displayed in Fig. 4.17. Here, the average and maximum size of the aggregates are shown as a function of the number of tasks performed by the cells, p, and the stickiness parameter σ. As one can see, in both models the group sizes increase with the stickiness. This result is expected since a higher stickiness implies a higher probability that the cells stay or come together, upon reproduction and aggregation. Note that the color gradation is presented in logarithmic scale as the evolved group sizes can vary significantly depending on the parameter values. The behavior of the group sizes with the number of tasks is more subtle. Remarkably, the size of the aggregates decreases with the number of tasks for the linear model case and increases for the spherical one. This outcome is observed both for average and maximum sizes. This means that, in the present model, only the spherical aggregates follow the size complexity rule and suggests that this rule is not universal but may present exceptions depending on factors like the life cycle and the aggregate geometry. Fig. 4.18 shows this phenomenon in greater detail, by focusing on the variation with the number of tasks. The effect of varying the number of tasks p is more pronounced for higher values of stickiness σ, especially in the case of spherical structures.

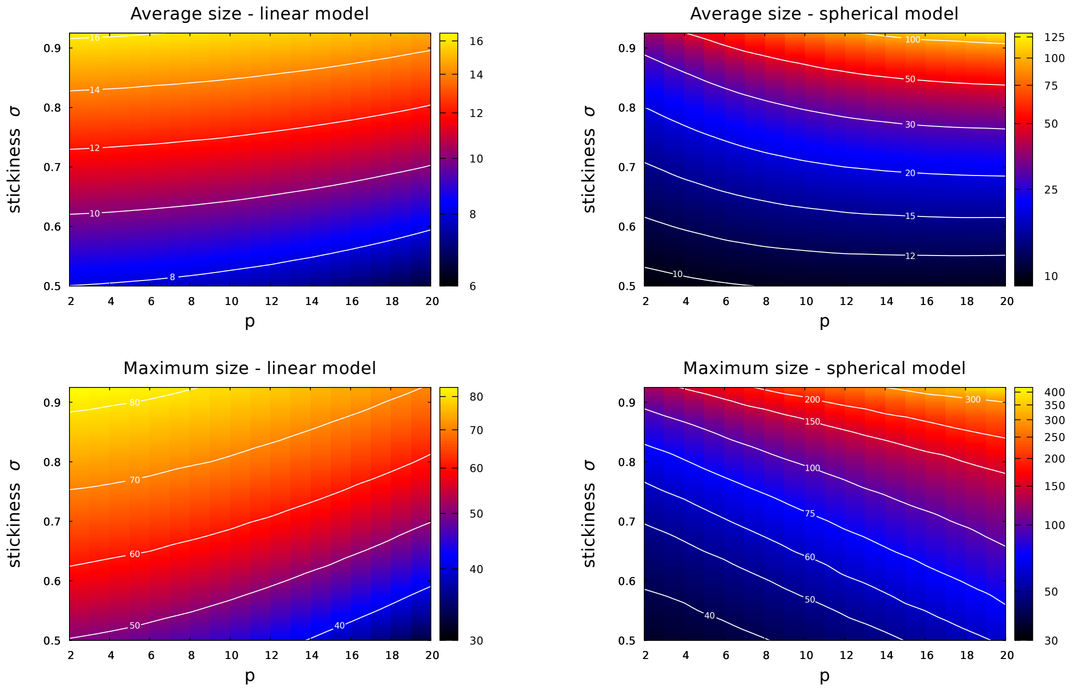

Figura 4.17: Average group size (upper panels) and maximum group size (lower panels) for the linear (left panels) and spherical models (right panels) in terms of the number of tasks p and stickiness σ. The remaining parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$, $k_D = 10^{-5}$ and $c = 0.04$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 8 independent runs. Figure adapted from [125].

It is important to ensure that this result is not only an artifact of the value of the cost parameter chosen. As such, we now explore the dependence of the

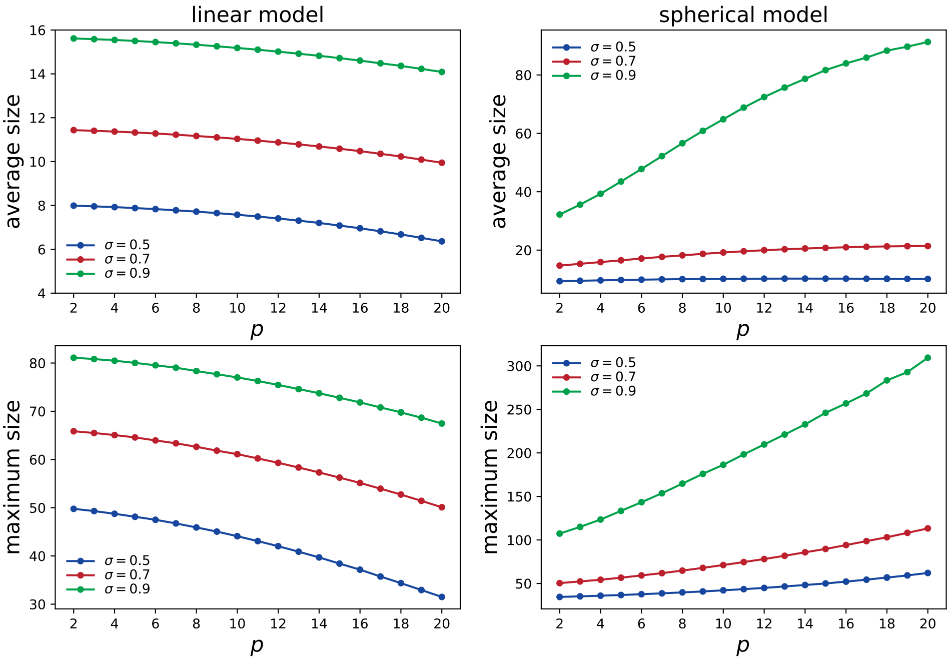

Figura 4.18: Average group size (upper panels) and maximum group size (lower panels) for the linear (left panels) and spherical models (right panels) as a function of the number of tasks p. The results are shown for the values of stickiness $\sigma = 0.5$ (blue) and 0.9 (red). The remaining parameter values are $k_+ = 0.01$, $k_- = 0.01$, $k_R = 0.01$, $k_D = 10^{-5}$ and $c = 0.04$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 100 independent runs. Figure adapted from [125].

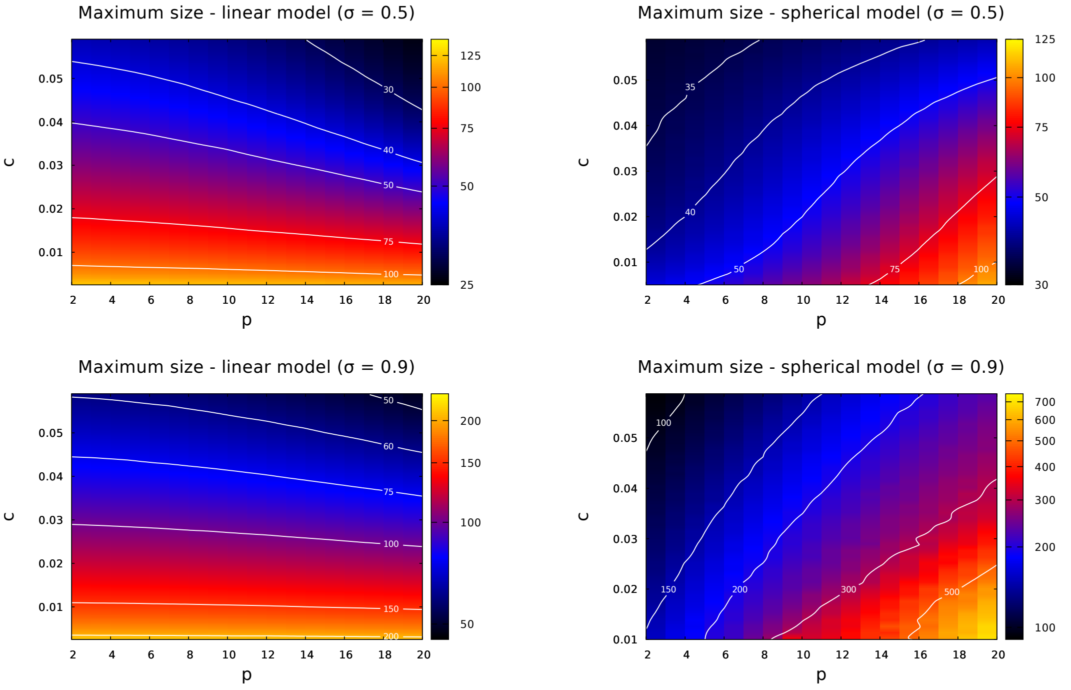

Figura 4.19: Maximum group size for the linear (left) and spherical (right) models in terms of the number of tasks p and cost parameter c. The remaining parameter values are $k_+ = 0$, $k_- = 0$, $k_R = 0.01$, $k_D = 10^{-5}$, $\sigma = 0.5$ (upper panels) and $\sigma = 0.9$ (lower panels). Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 8 independent runs. Figure adapted from [125].

number of groups on the cost parameter. Fig. 4.19 depicts the dependence of the maximum aggregate sizes with the cost parameter c and the number of tasks p, for two distinct values of stickiness σ. In general, one can notice that a higher cost c leads to smaller groups at equilibrium. The behavior is qualitatively the same regardless the value of σ, though a higher stickiness leads to larger groups, in agreement with previous outcomes in Fig. 4.17. These results corroborate our previous findings. Once again we see that the linear and spherical models display opposite behaviors as we vary the number of tasks p while keeping c constant, with the linear presenting a negative correlation between size and complexity p, and the spherical model showing a positive correlation.

4.3.3 Clonal development

Clonal development is arguably the most common mode of development adopted by multicellular life [126]. This regime is characterized by groups that are clones of one cell and do not receive members from other groups, staying together upon reproduction. As such, this mode of development can be modelled by suppressing the aggregation and dissociation mechanisms and restricting our analysis to high values of stickiness σ. In our analysis we will consider $k_+ = k_- = 0$ and set σ at 0.9, unless specified otherwise. Eliminating aggregation and focusing on development via staying together introduces some crucial advantages in relation to a "coming together" scheme. As all cells share a recent ancestor the genetic variation among cells of the organism is greatly reduced, eliminating much of the potential for evolutionary conflict [43, 127].

As Grosberg and Strathmann argue in [127] this is especially true in the case where a unicellular stage is included, since any genetic variation arising within the organism is redistributed among the offspring [127]. The distribution of genetic variation among the offspring provided by the unicellular stage brings two main types of benefits. First, it decreases the genetic competition between cells within the aggregate. Second, it provides a mechanism where the daughter groups can compete and eliminate deleterious mutations, that could otherwise accumulate if the group were to split without a unicellular stage. The benefits provided by these mechanisms seem to supersede the negative effects arising from the unicellular stage, e.g., an increased vulnerability to predation.

As in the clonal development mode we are constrained to a high stickiness regime, Fig. 4.20 displays the maximum aggregate sizes obtained under this regime in terms of the remaining free parameters, the cost parameter c and the number of tasks p. The figure exhibits not only the results for the linear and spherical models, but also for the snowflake-like structures. Snowflake structures present a much more chalenging computational problem since they require information of the whole structure to be stored and processed for each group independently. As such, they were only included in the context of clonal development. Some novel aspects surface in the analysis of the clonal development which were not included in the aggregative regime. The linear model, that previously presented negative correlation between the number of tasks p and the size, now includes a positive correlation for small p, followed by a negative correlation for large p, in the case of small values of the cost parameter c. For larger values of c, the previous behavior is recovered and only a negative correlation is found. Snowflake structures display a behavior qualitatively equivalent to the linear case, altough significantly larger structures are achieved. In its turn, the spherical model results repeat the pattern observed for the aggregative model. Over the full parameter range the aggregate sizes increase with the number of tasks p to be performed. In the limit of very low c

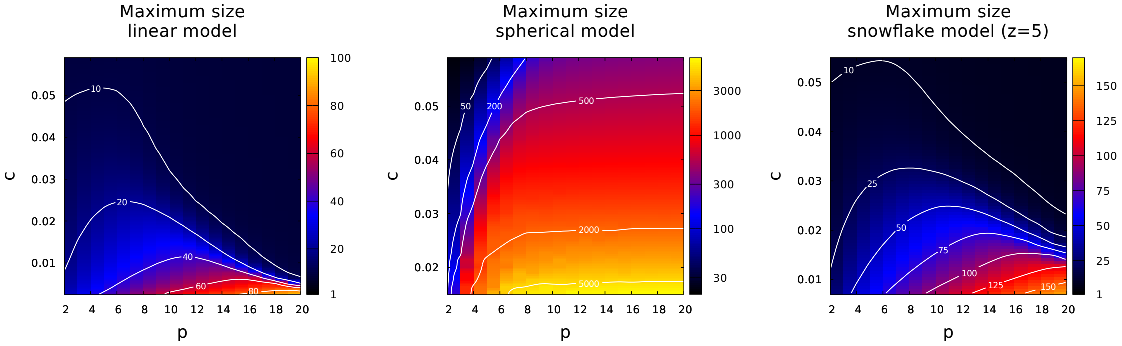

Figura 4.20: Maximum group size for linear (left panel), sphere-like (middle panel) and snowflake-like (right panel) structures as a function of the number of tasks p and cost parameter c. The remaining parameter values are $k_+ = 0$, $k_- = 0$, $k_R = 0.01$, $k_D = 10^{-5}$ and $\sigma = 0.9$. Each point corresponds to an average over 5×10^6 iterates sampled every 1000 iterates after reaching equilibrium for 8 independent runs. In the middle panel, the results are presented for $c \geq 0.015$. Figure adapted from [125].

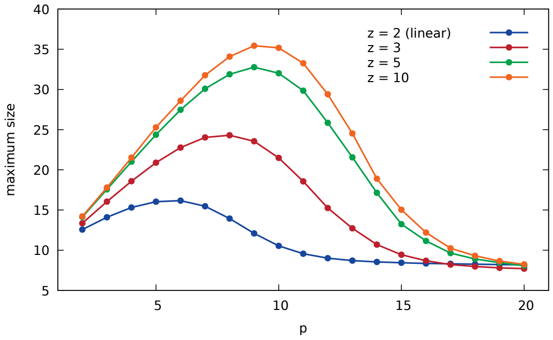

Figura 4.21: Maximum group size versus number of tasks p for snowflake-like structures. The parameter values are $c = 0.03$, $k_R = 0.01$, $k_D = 10^{-5}$ and stickiness $\sigma = 0.9$. In the clonal development mode $k_+ = k_- = 0$. Figure adapted from [125].

the group sizes grow significantly, e.g. larger than $7,000$ cells, rendering it difficult to explore this domain in detail. For this reason, the plot is constrained to values of $c \geq 0.015$.

Fig. 4.21 exhibits the way in which the morphology of snowflake-like structures affects the dynamics. The plot shows results for different values of maximum coordination number z, which develops a very important role. In an aggregate with maximum coordination number z, each cell can have up to z neighbors. This means that in a process of cell death, the number of new groups ranges from 1 to z. The sizes of these aggregates cannot be determined a priori since they depend on the location of the cell in the original aggregate. In the plot the cost c is set at $c = 0.03$. The case $z = 2$ corresponds to the linear model and shown in Fig. 4.20. When the maximum coordination z is increased, the peak of the maximum group size is shifted towards higher p, producing a larger region of p with a positive size-complexity correlation. Nevertheless, the position of the peak saturates around $p \approx 10$ and no appreciable difference is observed as z changes from 5 to 10. Besides producing a larger region where the size grows with p, a larger coordination number z also produces larger numbers, as hinted from Fig. 4.20.

4.4 Discussion and Conclusions

The model here proposed emerges from basic principles such as aggregation, dissociation, cell death and reproduction. As the groups can have an arbitrary number of cells, we need to further specify the group structure. The model can be used in conjunction with diverse group geometries. We explore linear chains, spherical aggregates and, in a later stage, include also snowflake-like structures. The reproduction rates are determined by a general fitness landscape that we assume to be maximized by the group at each moment. This fitness landscape results from tradeoffs between the different essential tasks that the organism should execute. A cell that simultaneously carries out several tasks incurs in a high fitness cost. The group formation can attenuate these costs by segregation of tasks, avoiding the main tradeoff costs while sharing the benefits from the tasks performed by the remaining cells in the group. The model relies on the assumption that changes in aggregate size allow the resulting aggregate to find new arrangements and explore new maxima of the fitness landscape, with direct implications on the cells' reproduction rates and opening room for the appearance of a division of labor.

Most approaches to the evolution of multicellularity pose the discussion either as a kin selection or a multilevel selection problem, in which the emergence of a higher level of biological organization is explained through the integration of cooperating individuals [88]. These higher level entities are favored by the fitness transfer from the lower level units to the higher level unit. In the mechanistic approach here proposed we model the mechanisms underlying the group formation directly, which allows us to probe their direct effect on the dynamics. This comes at the cost of losing the possibility of explicitly evaluating the fitness transfer from the lower level units to the higher level unit. Nevertheless, the resulting effect of larger group's fitness as agreggate size grows resembles the outcome of multilevel selection problems.

An alternative and elegant approach has been developed by Rueffler et al. [128]. In their approach, a performance is ascribed to each task, then the fitness is written as a monotonic increasing function of those performances. The tradeoffs are embodied in the performance functions. Following those assumptions, the authors explicitly find proper conditions that must be fulfilled by the system in order for specialization to evolve. Here the fitness does not only depend on the performance but also on the costs that arise from this performance, embedding the tradeoff in the definition of the fitness itself. Therefore, in our case, it is not possible to decouple and write the per capita reproductive rate of newly defined performance functions.

In a first step, we investigated the formation of groups required to perform two essential but poorly compatible tasks. The problem is addressed by means of extensive computer simulations that explore the relevant parameter space. All the processes can be depicted in the form of kinetic chemical reactions which allows us to define their corresponding reaction rates. This approach opens way to apply dynamic Monte-Carlo techniques to simulate the system dynamics, in our case the usage of a standard Gillespie algorithm [129]. Starting from a population comprised of single-celled organisms we observe the continuous formation of intermediate and large groups of cells, establishing a stable coexistence of aggregates of varying sizes. In spite of the underlying difference between group shapes, their qualitative behavior is not quite distinct with respect to the size of the largest groups produced and the size distribution at stationarity.

We observe that in both models a higher cost result in smaller groups and, as expected, the increase of stickiness favors the formation of larger aggregates. Although the cells are assumed to execute only two tasks, we still observe the formation of aggregates with more than two cells, implying an evolutionary advantage of larger groups. The higher robustness of larger groups, which experience smaller fitness changes upon variations of group size, probably contributes to this result.

A remarkable difference between the linear and spherical models is found when we analyze how group sizes evolve in terms of the aggregation and dissociation rates. For the linear model, the relation between group size and the rates of aggregation k_+ and dissociation k_- is straightforward: while the increase of k_+ favors increased group sizes, the augment of k_- favors smaller groups. Still, for the spherical model, the dependence of groups sizes on both k_+ and k_- is more subtle. At intermediate and large values of the aggregation rate, k_+, increased values of k_- can select for larger groups. In a recent paper, Ratcliff et al. found that, after some time of evolution, snowflakes structures developed division of labor with some cells undertaking programmed cell death (apoptosis) [50]. It is claimed that programmed cell death is evolutionary advantageous since it allows the snowflakes yeast to increase its fecundity. Indeed, they demonstrated that larger average group size and larger rates of apoptosis have coevolved [50]. In our model, dissociation plays the same role as cell death from the perspective of the group. Therefore, the spherical model, which allows smoother transitions in terms of variation in group size, seems to capture this feature of the empirical observation of the snowflakes yeast experiment [50].

We compare the situation where division of work is allowed with a situation where all the cells in the aggregate remain generalists to ascertain that our results depend on specialization mechanisms. We find that the division of work has always a measurable imprint on the system, being the most relevant in the limit of clonal lifecyle.

Our simulations show that group size evolves until it levels off and the mechanisms that favor group growth like aggregation and reproduction are balanced by the mechanisms supporting disruption. Particularly, it is worth mentioning that as time evolves in the linear chain model the average group size first passes through a peak and then drops to the stationary value. As aforementioned, aggregate sizes can grow further than two cells as an outcome of fitness reshaping which, in its turn, allows the whole organism to explore and find higher levels of fitness, i.e., new arrangements are found in which the process of division of labor can turn even more effective. One may also conjecture that the maximum sizes reached in this dynamics reflect an upper bound for aggregate sizes, and a further growth in the number of cells comprising the organisms requires additional elements, such as the execution of new tasks beneficial to the group, or other strategies that enable the transfer of fitness from the individual level to the group level. In brief, from this point on, more division of labor will be required.

After analysing the two tasks model we focused our attention to a more general approach which permits higher levels of division of labor. Another essential characteristic of this new iteration of the model is that it now allows us to tune the complexity of the system by adjusting the number of functions that should be performed by the aggregates.

Our findings reveal that the shape and topology of the aggregates play a major role in the evolutionary dynamics of group formation. The mode of body formation can also lead to quite distinct outcomes. Compact structures, like the ones generated by the spherical model, seem to more strictly follow the size-complexity rule, which conjectures that aggregate sizes positively correlate with the number of specialized functions performed by the multicellular organism. Surprisingly, we observe that the linear structure clearly violates this conjecture, especially for the aggregative mode of formation, where a negative correlation between size and number of tasks is found in all the domain of the parameter space. For the clonal mode of development, the scenario is more subtle and one observes a positive correlation between size and number of tasks at small values of the cost parameter c, and a less trivial dependence at intermediate values of c. For very large c, the decrease of aggregate size with number of tasks is recovered.

These outcomes evince how the biological mechanisms interact with the geometric constraints. Particularly, whereas in compact structures the mechanisms of group division (cell death and dissociation) lead to the removal of a single cell in the original aggregate, in linear structures those mechanisms can result in aggregates of largely reduced sizes, thus leading to big variation in the mean fitness. The interplay between cell death (apoptosis) and geometry constraints has been observed in experiments with yeast populations under gravity selection [50, 130], where yeast lineages evolved snowflake-like structures. In spite of a poor performance with respect to gravity selection, selection for a higher rate of apoptosis was ascertained.

The snowflake-like structures, as observed in the experiment mentioned above, are also addressed here. This allows us to better explore the role of structure and topology. However, its study is restricted to the clonal mode of development mainly due to modeling limitations. Snowflake-like structures reveal a dual character and interpolate between the linear model and the spherical model. At small and even intermediate levels of cost c, a positive correlation between aggregate size and the number of tasks p is seen, and from intermediate to large costs there is a bell-shaped dependence of aggregate size on p. Although the augment of the maximum coordination number z can make the structure tighter, its effect saturates and the limit of large z does not replicate the outcomes of sphere-like structures. The main reason is that in a link of snowflake-like structure with coordination number z_n, cell death gives rise to $z_n - 1$ new aggregates of variable sizes which depend on how those branches developed, a behaviour quite different to the one witnessed in spherical structures. However, it is important to highlight that if the number of tasks a cell can develop is limited, then for a wide range of parameter values the dependence of aggregate size on the number of tasks complies with the size-complexity rule, as can be seen in Fig. 5. According to the results presented by Bell and Bonner in Refs. [117, 118], the set of organisms embraced by this limit is substantial. In their studies, organisms belonging to different phyla were put together in order to analyze the relationship between size and complexity.

As future perspectives, we believe that the current approach can open new lines in the debate about the evolution of complexity and can be extended to become more realistic. For instance, we have not explicitly addressed the emergence of germ-soma differentiation, focusing instead on tradeoffs between generic tasks in the organism. This is a relevant question since the separation between somatic and germ cell lines is considered to be one of the earlier differentiations in the evolution of many multicellular lineages [131]. Gavrilets [96] and Solari et al. [132] provide interesting theoretical approaches to the evolution of this type of differentiation. Both studies are inspired in the biology of volvocine green algae, which features compact structures and a positive correlation between the colony size and differentiation level [133]. This is in consistent with our own findings since compact structures follow the size-complexity rule. In Gavrilets model, colony size plays a minor role on the dynamics, limiting its direct application to the context of the size-complexity rule. On the other hand, the work of Solari et al. [132] allows us to assess the effect of aggregate size on differentiation. They assume that the compact spherical geometry of the aggregates introduces growing costs with size owing to nutrient influx limitations. These growing costs should then be overcome either through the appearance of specialized cells or adoption of different geometries, not entailing increased costs with size. Nevertheless, they do not explicitly model the effect of geometry in group reproduction, that in our approach naturally appears through group division mechanisms. Since here the topology of the system plays an important role in the evolution of multicellular aggregates, it is also of interest to explore in the near future the role of geometry in systems that explicitly takes into account the viability-fecundity tradeoff. Other aspects can also be taken into account, such as size dependent death rates that appear, for instance, due to preferential predation of smaller organisms. This is though to be a relevant drive for the emergence of undifferentiated multicellular groups in the first place. A size dependent death rate is included also in the experimental work by Ratcliff and collaborators [50, 51], where gravity is used to select for larger groups.

In short, the framework here developed enabled us to explore the relationship between organism size and the number of potential tasks that can be performed by a cell. Our findings suggest that the size-complexity rule, as proposed by Bell and

Bonner [117, 118] may not always be valid. A more delicate relationship between size and complexity arises, that can be affected by aspects such as the topology of the groups and the developmental mode. Although under certain conditions a positive correlation between size and complexity appears, we cannot expect this relation to hold independently of the details of the system under analysis. Particularly, it has been observed that more compact structures, like spherical structures, are more prone to follow the claim of the size-complexity rule. On the other hand, less robust structures such as linear chains, which are more vulnerable to drastic changes due to division mechanisms, can, in a broad domain of the parameter space, violate the size-complexity rule.

The main results presented in this chapter have been published in two articles:

- Ref. [125]: A theoretical approach to the size-complexity rule, André Amado, Carlos Batista, Paulo R. A. Campos, Evolution (2017)
  DOI: 10.1111/evo.13392

- Ref. [122]: A mechanistic model for the evolution of multicellularity, André Amado, Carlos Batista, Paulo R. A. Campos, Physica A (2018)
  DOI: 10.1016/j.physa.2017.11.080

The code used to perform the numerical simulations is available online in Dryad repository at

- Ref. [134]: Data from: A theoretical approach to the size-complexity rule. André Amado, Carlos Batista, Paulo R. A. Campos, Dryad Digital Repository (2017)
  DOI: 10.5061/dryad.j26k3

5 CONCLUSION

In this work, we have proposed and analyzed three models that address the establishment of multicellularity and the development of biological complexity. This is an extremely interesting area of work, recently revived by a number of experiments ranging from the massive genetic mapping of species to the experimental evolution of multicellularity in laboratorial conditions.

As a first step, we have studied a model that focus on the competition between strains with different metabolic properties in the context of structured and well-mixed populations. We explore the fact that there is a tradeoff between the rate of resource acquisition and the yield of resource processing. The model is a multilevel model, adequate to the analysis of the transition between the transition from a unicellular life style to a multicellular one. The results suggest that an association between the appearance of multicellularity and a more efficient metabolism since the more efficient strain is favored over a much wider range of parameters in the case of structured populations comparing with the well-mixed scenario. This is consistent with the theories about the development of multicellularity that place the origin of multicellularity at roughly the same time that the oxygen levels rose in the planet and allowed the generalization of the much more efficient aerobic respiration method of metabolism. There is a large parameter region where the social conflict is present and the inneficient strain dominated inside the group, but a pure group of efficient cells outcompetes a group of inefficient ones, opening a window for the fixation of the efficient strain in the population.

In the sequence, we investigated a model for a simple multicellular system subject to tradeoffs. The organism is subject to mutation and two levels of selection, specifically fertility and viability selection. For the purpose of this work, fertility has been kept constant, but a generalized situation where it is also subject to mutation is worth studying, since fertility rate is one essential life history trait. The model considers essential tasks that are performed by the organism. Viability is defined by the performance of the organism in different tasks. The execution of these tasks is controlled at the levels of structural and regulatory genes. Structural genes are responsible by defining the output of the function directly, while regulatory genes control the level of expression of the function. By suppressing some functions in specific cell types, the organism can isolate incompatible tasks to different cells effectively curbing the effect of the tradeoff. We studied the cell differentiation process in terms of different abstract scenarios for the tradeoffs between tasks.

As a concrete illustration, we applied the model to a system of cyanobacteria. Cyanobacteria need both a carbon and a nitrogen source to survive. Yet, the carbon and nitrogen fixation processes present a strong tradeoff since the fixation of oxygen produces molecular oxygen O_2 which damages the enzyme system responsible by the fixation of nitrogen. The model reproduced the behavior observe in nature, where cells with low access to an external source of biologically usable nitrogen developed cell differentiation between generalist cells and cells specialized in nitrogen fixation (called heterocysts). In contrast, in a scenario of abundant nitrogen supply no specialization took place.

The last chapter is devoted to relationship between size and complexity. This relationship has been the subject of intense study. Formally introduced in the Biological context by Bell and Mooers [117], the size-complexity rule claims that there is a direct relation between the size of an organism and its complexity. For this purpose, the complexity of an organism is measured in terms the number of different cell types that it displays. We propose and study a mechanistic model for the dynamics of multicellular aggregates and finally apply it to this problem. The model incorporates the mechanisms of cell reproduction, cell death, aggregation and dissociation. We consider three possible aggregate geometries: a compact sphere-like geometry, a linear geometry and a snowflake-like geometry. The reproduction rate depends on a number of somatic tasks that display tradeoffs, and are considered to take their optimal value given the conditions. We conclude that compact aggregates follow the size-complexity rule, whereas more fragile aggregates, like linear or snowflake-like structures, display a more complex behavior. When the organism has a number of tasks to perform there is direct relationship between size and complexity. However, for a large number of tasks, this relationship inverts and a higher number of tasks leads to smaller-sized organisms.

The works here discussed have the potential to be generalized in a number of interesting ways, as described in their sectorial conclusions.

REFERENCES

[1] Thomas R. Malthus. *An essay on the principle of population*. 1798.

[2] Charles Darwin and Alfred Wallace. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society*, 3(9):45–62, 1858.

[3] Charles Darwin. *On the origin of the species by means of natural selection*. John Murray, 1859.

[4] Ronald Fisher. XV. — The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.

[5] Claude J. Allègre, Gérard Manhès, and Christa Göpel. The age of the Earth. *Geochimica et Cosmochimica Acta*, 59(8):1445 – 1456, 1995.

[6] G. Brent Dalrymple. The age of the Earth in the twentieth century: a problem (mostly) solved. *Geological Society, London, Special Publications*, 190(1):205–221, 2001.

[7] Claude J. Allègre, Gérard Manhès, and Christa Göpel. The major differentiation of the Earth at 4.45 Ga. *Earth and Planetary Science Letters*, 267(1–2):386 – 398, 2008.

[8] Alex N. Halliday. The origin of the Earth - what's new? *Elements*, 2(4):205–210, 2007.

[9] Jacques Laskar, Frédéric Joutel, and Philippe Robutel. Stabilization of the Earth's obliquity by the Moon. *Nature*, 361(6413):615–617, 1993.

[10] Ashwini K. Lal. Origin of life. *Astrophysics and Space Science*, 317(3):267–278, 2008.

[11] Simon A. Wilde, John W. Valley, William H. Peck, and Colin M. Graham. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature*, 409(6817):175–178, 2001.

[12] John A. Tarduno, Rory D. Cottrell, William J. Davis, Francis Nimmo, and Richard K. Bono. A Hadean to Paleoarchean geodynamo recorded by single zircon crystals. *Science*, 349(6247):521–524, 2015.

[13] Leslie E. Orgel. The origin of life — a review of facts and speculations. *Trends in biochemical sciences*, 23(12):491–495, 1998.

[14] Yecheskel Wolman, William J. Haverland, and Stanley L. Miller. Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. *Proceedings of the National Academy of Sciences*, 69(4):809–811, 1972.

[15] Christopher F. Chyba, Paul J. Thomas, Leigh Brookshaw, and Carl Sagan. Cometary delivery of organic molecules to the early Earth. *Science*, 249(4967):366–374, 1990.

[16] Christopher Chyba and Carl Sagan. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature*, 355(6356):125, 1992.

[17] Pascale Ehrenfreund and Jan Cami. Cosmic carbon chemistry: from the interstellar medium to the early Earth. *Cold Spring Harbor perspectives in biology*, 2(12):a002097, 2010.

[18] Martin Ferus, David Nesvornỳ, Jiří Šponer, Petr Kubelík, Regina Michalčíková, Violetta Shestivská, Judit E. Šponer, and Svatopluk Civiš. High-energy chemistry of formamide: A unified mechanism of nucleobase formation. *Proceedings of the National Academy of Sciences*, 112(3):657–662, 2015.

[19] J. William Schopf. Microfossils of the early archean apex chert: New evidence of the antiquity of life. *Science*, 260(5108):640–646, 1993.

[20] Manfred Schidlowski. A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature*, 333(6171):313–318, 1988.

[21] Elizabeth A. Bell, Patrick Boehnke, T. Mark Harrison, and Wendy L. Mao. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proceedings of the National Academy of Sciences*, 112(47):14518–14521, 2015.

[22] Madeline C. Weiss, Filipa L. Sousa, Natalia Mrnjavac, Sinje Neukirchen, Mayo Roettger, Shijulal Nelson-Sathi, and William F. Martin. The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, 1:16116, 2016.

[23] Johann P. Gogarten and David Deamer. Is LUCA a thermophilic progenote? *Nature Microbiology*, 1:16229, 2016.

[24] Carl R. Woese, Otto Kandler, and Mark L. Wheelis. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12):4576–4579, 1990.

[25] James R. Brown and W. Ford Doolittle. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proceedings of the National Academy of Sciences*, 92(7):2441–2445, 1995.

[26] James R. Brown, Christophe J. Douady, Michael J. Italia, William E. Marshall, and Michael J. Stanhope. Universal trees based on large combined protein sequence data sets. *Nature genetics*, 28(3):281–285, 2001.

[27] Tom A. Williams, Peter G. Foster, Cymon J. Cox, and T. Martin Embley. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, 504(7479):231–236, 2013.

[28] Anja Spang, Jimmy H. Saw, Steffen L. Jørgensen, Katarzyna Zaremba-Niedzwiedzka, Joran Martijn, Anders E. Lind, Roel van Eijk, Christa Schleper, Lionel Guy, and Thijs J. G. Ettema. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551):173–179, 2015.

[29] Patrick Forterre. The origin of viruses and their possible roles in major evolutionary transitions. *Virus research*, 117(1):5–16, 2006.

[30] Patrick Forterre. Defining life: The virus viewpoint. *Origins of Life and Evolution of Biospheres*, 40(2):151–160, Apr 2010.

[31] Carlos Canchaya, Ghislain Fournous, Sandra Chibani-Chennoufi, Marie-Lise Dillmann, and Harald Brüssow. Phage as agents of lateral gene transfer. *Current opinion in microbiology*, 6(4):417–424, 2003.

[32] Laura S. Frost, Raphael Leplae, Anne O. Summers, and Ariane Toussaint. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, 3(9):722–732, 2005.

[33] J. Cameron Thrash, Alex Boyd, Megan J. Huggett, Jana Grote, Paul Carini, Ryan J. Yoder, Barbara Robbertse, Joseph W. Spatafora, Michael S. Rappé, and Stephen J. Giovannoni. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Scientific reports*, 1:13, 2011.

[34] Eörs Szathmáry and John Maynard Smith. The major evolutionary transitions. *Nature*, 374(6519):227–232, 1995.

[35] John Maynard Smith and Eors Szathmary. *The major transitions in evolution*. Oxford University Press, 1997.

[36] Richard E. Michod and Denis Roze. Transitions in individuality. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1383):853–857, 1997.

[37] Richard E. Michod and Aurora M. Nedelcu. On the reorganization of fitness during evolutionary transitions in individuality. *Integrative and Comparative Biology*, 43(1):64–73, 2003.

[38] William M. Wheeler. The ant-colony as an organism. *Journal of Morphology*, 22(2):307–325, 1911.

[39] Martin A. Nowak, Corina E. Tarnita, and Edward O. Wilson. The evolution of eusociality. *Nature*, 466(7310):1057–1062, 2010.

[40] Michael W. Gray, Gertraud Burger, and B. Franz Lang. Mitochondrial evolution. *Science*, 283(5407):1476–1481, 1999.

[41] Geoffrey I. McFadden. Chloroplast origin and integration. *Plant Physiology*, 125(1):50–53, 2001.

[42] Takuro Nakayama and John M. Archibald. Evolving a photosynthetic organelle. *BMC biology*, 10(1):35, 2012.

[43] Richard K. Grosberg and Richard R. Strathmann. The evolution of multicellularity: a minor major transition? *Annual Review of Ecology, Evolution, and Systematics*, pages 621–654, 2007.

[44] Karl J. Niklas. The evolutionary-developmental origins of multicellularity. *American Journal of Botany*, 101(1):6–25, 2014.

[45] James A. Shapiro. Bacteria as multicellular organisms. *Scientific American*, 258(6):82–89, 1988.

[46] James A. Shapiro. Thinking about bacterial populations as multicellular organisms. *Annual Review of Microbiology*, 52(1):81–104, 1998. PMID: 9891794.

[47] Jeremy S. Webb, Michael Givskov, and Staffan Kjelleberg. Bacterial biofilms: prokaryotic adventures in multicellularity. *Current opinion in microbiology*, 6(6):578–585, 2003.

[48] Dennis Claessen, Daniel E. Rozen, Oscar P. Kuipers, Lotte Søgaard-Andersen, and Gilles P. Van Wezel. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nature Reviews Microbiology*, 12(2):115–124, 2014.

[49] Enrique Flores and Antonia Herrero. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology*, 8(1):39–50, 2010.

[50] William C. Ratcliff, R. Ford Denison, Mark Borrello, and Michael Travisano. Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, 109(5):1595–1600, 2012.

[51] William C. Ratcliff, Matthew D. Herron, Kathryn Howell, Jennifer T. Pentz, Frank Rosenzweig, and Michael Travisano. Experimental evolution of an alternating uni-and multicellular life cycle in Chlamydomonas reinhardtii. *Nature communications*, 4, 2013.

[52] Yuriy Pichugin, Jorge Peña, Paul B. Rainey, and Arne Traulsen. Fragmentation modes and the evolution of life cycles. *PLOS Computational Biology*, 13(11):1–20, 11 2017.

[53] William D. Hamilton. The evolution of altruistic behavior. *The American Naturalist*, 97(896):354–356, 1963.

[54] Robert Axelrod and William D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[55] Robert Axelrod. *The evolution of cooperation*, volume 5145. Basic Books (AZ), 1984.

[56] Martin A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.

[57] Richard Law. Optimal life histories under age-specific predation. *The American Naturalist*, 114(3):399–417, 1979.

[58] Martin A. Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[59] Thomas Pfeiffer, Stefan Schuster, and Sebastian Bonhoeffer. Cooperation and competition in the evolution of ATP-producing pathways. *Science*, 292(5516):504–507, 2001.

[60] Donald E. Canfield. The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annual Review of Earth and Planetary Sciences*, 33:1–36, 2005.

[61] Bettina E. Schirrmeister, Jurriaan M. de Vos, Alexandre Antonelli, and Homayoun C. Bagheri. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences*, 110(5):1791–1796, 2013.

[62] S. Blair Hedges, Jaime E. Blair, Maria L. Venturi, and Jason L. Shoe. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology*, 4(1):2, 2004.

[63] Fernando Angulo-Brown, Moisés Santillán, and E. Calleja-Quevedo. Thermodynamic optimality in some biochemical reactions. *Il Nuovo Cimento D*, 17(1):87–90, 1995.

[64] Keith L. Manchester. Optimization of energy coupling: what is all the argument about? *Biochemistry and Molecular Biology Education*, 28(1):18–19, 2000.

[65] Maja Novak, Thomas Pfeiffer, Richard E. Lenski, Uwe Sauer, and Sebastian Bonhoeffer. Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli. *The American Naturalist*, 168(2):242–251, 2006.

[66] R. Craig MacLean and Ivana Gudelj. Resource competition and social conflict in experimental populations of yeast. *Nature*, 441(7092):498–501, 2006.

[67] R. Craig MacLean. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity*, 100(3):233–239, 2008.

[68] David C. Catling, Christopher R. Glein, Kevin J. Zahnle, and Christopher P. McKay. Why O_2 is required by complex life on habitable planets and the concept of planetary ”oxygenation time”. *Astrobiology*, 5(3):415–438, 2005.

[69] Claudia Acquisti, Jürgen Kleffe, and Sinéad Collins. Oxygen content of transmembrane proteins over macroevolutionary time scales. *Nature*, 445(7123):47–52, 2007.

[70] Victor J. Thannickal. Oxygen in the evolution of complex life and the price we pay. *American Journal of Respiratory Cell and Molecular Biology*, 40(5):507–510, 2009.

[71] Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.

[72] James R. Broach. Nutritional control of growth and development in yeast. *Genetics*, 192(1):73–105, 2012.

[73] Kazuyuki Shimizu. Metabolic regulation of a bacterial cell system with emphasis on Escherichia coli metabolism. *ISRN biochemistry*, 2013, 2013.

[74] Arthur L. Koch. Microbial physiology and ecology of slow growth. *Microbiology and Molecular Biology Reviews*, 61(3):305–318, 1997.

[75] Filip Rolland, Joris Winderickx, and Johan M. Thevelein. Glucose-sensing and-signalling mechanisms in yeast. *FEMS yeast research*, 2(2):183–201, 2002.

[76] Lucinda Notley-McRobb, Alison Death, and Thomas Ferenci. The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology*, 143(6):1909–1918, 1997.

[77] Eunüs S. Ali, Jin Hua, Claire H. Wilson, George A. Tallis, Fiona H. Zhou, Grigori Y. Rychkov, and Greg J. Barritt. The glucagon-like peptide-1 analogue exendin-4 reverses impaired intracellular Ca^{2+} signalling in steatotic hepatocytes. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1863(9):2135–2146, 2016.

[78] Lenin Fernández and Paulo R. A. Campos. Evolution of complexity in a resource-based model. *The European Physical Journal Plus*, 132(2):72, 2017.

[79] André Amado, Lenin Fernández, Weinei Huang, Fernando F. Ferreira, and Paulo R. A. Campos. Competing metabolic strategies in a multilevel selection model. *Royal Society Open Science*, 3:160544, 2016.

[80] Tobias Frick and Stefan Schuster. An example of the prisoner's dilemma in biochemistry. *Naturwissenschaften*, 90(7):327–331, 2003.

[81] Karin Otterstedt, Christer Larsson, Roslyn M. Bill, Anders Ståhlberg, Eckhard Boles, Stefan Hohmann, and Lena Gustafsson. Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. *EMBO reports*, 5(5):532–537, 2004.

[82] André Amado, Lenin Fernández, Weini Huang, Fernando F. Ferreira, and Paulo R. A. Campos. Data from: Competing metabolic strategies in a multilevel selection model, 2016.

[83] Sewall Wright. Isolation by distance. *Genetics*, 28(2):114, 1943.

[84] Matthew G. Vander Heiden, Lewis C. Cantley, and Craig B. Thompson. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science*, 324(5930):1029–1033, 2009.

[85] Lenin Fernández, André Amado, Paulo R. A. Campos, and Fernando Fagundes Ferreira. Coexistence of competing metabolic pathways in well-mixed populations. *Physical Review E*, 93:052401, 2016.

[86] Yoriko Saeki, Midori Tuda, and Philip H. Crowley. Allocation tradeoffs and life histories: a conceptual and graphical framework. *Oikos*, 123(7):786–793, 2014.

[87] Amy L. Angert, Travis E. Huxman, Peter Chesson, and D. Lawrence Venable. Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences*, 106(28):11641–11645, 2009.

[88] Richard E. Michod. Evolution of individuality during the transition from unicellular to multicellular life. *Proceedings of the National Academy of Sciences*, 104(suppl 1):8613–8618, 2007.

[89] Helen Olofsson, Jörgen Ripa, and Niclas Jonzén. Bet-hedging as an evolutionary game: the trade-off between egg size and number. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20090500, 2009.

[90] Matjaž Perc and Paolo Grigolini. Collective behavior and evolutionary games–An introduction. *Chaos, Solitons & Fractals*, 56:1–5, 2013.

[91] Attila Szolnoki and Matjaž Perc. Group-size effects on the evolution of cooperation in the spatial public goods game. *Physical Review E*, 84(4):047102, 2011.

[92] Marco A. Javarone. Statistical physics of the spatial Prisoner's Dilemma with memory-aware agents. *The European Physical Journal B*, 89(2):42, 2016.

[93] Bjørn Østman, Randall Lin, and Christoph Adami. Trade-offs drive resource specialization and the gradual establishment of ecotypes. *BMC Evolutionary Biology*, 14(1):1, 2014.

[94] Yael Artzy-Randrup and Alexey S. Kondrashov. Sympatric speciation under incompatibility selection. *Proceedings of the National Academy of Sciences*, 103(31):11619–11624, 2006.

[95] Jacqueline Le Vinh Thuy, John M. VandenBrooks, and Michael Angilletta. Developmental plasticity evolved according to specialist-generalist trade-offs in experimental populations of Drosophila melanogaster. *Biology Letters*, 12(7), 2016.

[96] Sergey Gavrilets. Rapid transition towards the division of labor via evolution of developmental plasticity. *PLOS Computational Biology*, 6(6):e1000805, 2010.

[97] Ivana Gudelj, Robert Beardmore, Sinan S. Arkin, and R. Craig MacLean. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. *Journal of Evolutionary Biology*, 20(5):1882–1889, 2007.

[98] Justin R. Meyer, Ivana Gudelj, and Robert Beardmore. Biophysical mechanisms that maintain biodiversity through trade-offs. *Nature communications*, 6, 2015.

[99] David Tilman. Competition and biodiversity in spatially structured habitats. *Ecology*, 75(1):2–16, 1994.

[100] Derek A. Roff, Serge Mostowy, and Daphne J. Fairbairn. The evolution of trade-offs: testing predictions on response to selection and environmental variation. *Evolution*, 56(1):84–95, 2002.

[101] Alison E Bennett, Janice Alers-Garcia, and James D Bever. Three-way interactions among mutualistic mycorrhizal fungi, plants, and plant enemies: hypotheses and synthesis. *The American Naturalist*, 167(2):141–152, 2005.

[102] Lesley T. LANCASTER, Lisa C. HAZARD, Jean CLOBERT, and Barry R. SINERVO. Corticosterone manipulation reveals differences in hierarchical organization of multidimensional reproductive trade-offs in r-strategist and k-strategist females. *Journal of Evolutionary Biology*, 21(2):556–565, 2008.

[103] Kyle F. Edwards and John J. Stachowicz. Multivariate trade-offs, succession, and phenological differentiation in a guild of colonial invertebrates. *Ecology*, 91(11):3146–3152, 2010.

[104] Kyle F Edwards, Christopher A Klausmeier, and Elena Litchman. Evidence for a three-way trade-off between nitrogen and phosphorus competitive abilities and cell size in phytoplankton. *Ecology*, 92(11):2085–2095, 2011.

[105] María Uriarte, James S. Clark, Jess K. Zimmerman, Liza S. Comita, Jimena Forero-Montaña, and Jill Thompson. Multidimensional trade-offs in species responses to disturbance: implications for diversity in a subtropical forest. *Ecology*, 93(1):191–205, 2012.

[106] Salva Duran-Nebreda and Ricard Solé. Emergence of multicellularity in a model of cell growth, death and aggregation under size-dependent selection. *Journal of The Royal Society Interface*, 12(102):20140982, 2015.

[107] André Amado and Paulo R. A. Campos. The influence of the composition of tradeoffs on the generation of differentiated cells. *Journal of Statistical Mechanics*, page 063501, 2017.

[108] Raymond J. H. Beverton and Sidney J. Holt. *On the dynamics of exploited fish populations*, volume 11. Springer Science & Business Media, 2012.

[109] Iaroslav Ispolatov, Martin Ackermann, and Michael Doebeli. Division of labour and the evolution of multicellularity. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1734):1768–1776, 2012.

[110] Lai-Wa Tam and David L. Kirk. The program for cellular differentiation in Volvox carteri as revealed by molecular analysis of development in a gonidialess/somatic regenerator mutant. *Development*, 112(2):571–580, 1991.

[111] Peter H. Van Tienderen. Evolution of generalists and specialist in spatially heterogeneous environments. *Evolution*, pages 1317–1331, 1991.

[112] Valentina Rossetti and Homayoun C. Bagheri. Advantages of the division of labour for the long-term population dynamics of cyanobacteria at different latitudes. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1742):3457–3466, 2012.

[113] Birgitta Bergman, John R. Gallon, Amar N. Rai, and Lucas J. Stal. N_2 fixation by non-heterocystous cyanobacteria. *FEMS Microbiology Reviews*, 19(3):139, 1997.

[114] Bettina E. Schirrmeister, Alexandre Antonelli, and Homayoun C. Bagheri. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology*, 11(1):45, 2011.

[115] Pavel L. Krapivsky, Sidney Redner, and Eli Ben-Naim. *A kinetic view of statistical physics*. Cambridge University Press, 2010.

[116] Shay Gueron and Simon A. Levin. The dynamics of group formation. *Mathematical Biosciences*, 128(1):243–264, 1995.

[117] Graham Bell and Arne O. Mooers. Size and complexity among multicellular organisms. *Biological Journal of the Linnean Society*, 60(3):345–363, 1997.

[118] John T. Bonner. Perspective: the size-complexity rule. *Evolution*, 58(9):1883–1890, 2004.

[119] Daniel W. McShea. Functional complexity in organisms: parts as proxies. *Biology and Philosophy*, 15(5):641–668, 2000.

[120] Robert L. Carneiro. On the relationship between size of population and complexity of social organization. *Southwestern Journal of Anthropology*, 23(3):234–243, 1967.

[121] Rudolf Clausius. XVI. On a mechanical theorem applicable to heat. *Philosophical Magazine*, 40(265):122–127, 1870.

[122] André Amado, Carlos Batista, and Paulo R. A. Campos. A mechanistic model for the evolution of multicellularity. *Physica A: Statistical Mechanics and its Applications*, 492C:1543–1554, 2018.

[123] Corina E. Tarnita, Clifford H. Taubes, and Martin A. Nowak. Evolutionary construction by staying together and coming together. *Journal of Theoretical Biology*, 320:10–22, 2013.

[124] Charles D. Michener. *The social behavior of the bees: a comparative study*, volume 73. Harvard University Press, 1974.

[125] André Amado, Carlos Batista, and Paulo R. A. Campos. A theoretical approach to the size-complexity rule. *Evolution*, 72(1):18–29, 2018.

[126] John T. Bonner. The origins of multicellularity. *Integrative Biology Issues News and Reviews*, 1(1):27–36, 1998.

[127] Richard K. Grosberg and Richard R. Strathmann. One cell, two cell, red cell, blue cell: the persistence of a unicellular stage in multicellular life histories. *Trends in ecology & evolution*, 13(3):112–116, 1998.

[128] Claus Rueffler, Joachim Hermisson, and Günter P. Wagner. Evolution of functional specialization and division of labor. *Proceedings of the National Academy of Sciences*, 109(6):E326–E335, 2012.

[129] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[130] Eric Libby, William Ratcliff, Michael Travisano, and Ben Kerr. Geometry shapes evolution of early multicellularity. *PLOS Computational Biology*, 10(9):e1003803, 2014.

[131] Carl Simpson. The evolutionary history of division of labour. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20110766, 2011.

[132] Cristian A. Solari, John O. Kessler, and Raymond E. Goldstein. A general allometric and life-history model for cellular differentiation in the transition to multicellularity. *The American Naturalist*, 181(3):369–380, 2013.

[133] David L. Kirk. *Volvox: a search for the molecular and genetic origins of multicellularity and cellular differentiation*, volume 33. Cambridge University Press, 2005.

[134] A Amado, C Batista, and PRA Campos. Data from: A theoretical approach to the size-complexity rule, 2017.

[135] Patrick A. P. Moran. Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 60–71. Cambridge University Press, 1958.

[136] Joseph L. Doob. Topics in the theory of Markoff chains. *Transactions of the American Mathematical Society*, 52(1):37–64, 1942.

[137] Joseph L. Doob. Markoff chains–denumerable case. *Transactions of the American Mathematical Society*, 58(3):455–473, 1945.

[138] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

APPENDIX A – LINEAR STABILITY OF EQUILIBRIUM SOLUTIONS

Many time dependent equations allow for equilibrium solutions, where the solution becomes independent of time. These solutions are important since frequently they represent the state of the system in the long time limit. Often the system starts out of the equilibrium and evolves to an equilibrium after an initial transient period when the solution depends on time. When an equilibrium solution is found it should be determine whether this solution is stable or not, i.e., if a small perturbation is applied to the system does it remain in the same state or moves out? The simplest method to determine this is the linear stability analysis. This analysis determines if an infinitesimal perturbation to the solution will grow or be damped.

The linear stability analysis of equilibrium solutions works differently on continuous time and discrete time equations. We will start by describing the linear stability of solutions of continuous time differential equations. Given a set of n differential equations in the form

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}t} &= f_1(t, y_1, ..., y_n)\\ &\vdots\\ \frac{\mathrm{d}y_n}{\mathrm{d}t} &= f_n(t, y_1, ..., y_n),\end{aligned} \tag{A.1}$$

the equilibrium solution $\hat{\mathbf{y}} \equiv (\hat{y}_1, \ldots, \hat{y}_n)$ is such that $\left.\frac{\mathrm{d}y_i}{\mathrm{d}t}\right|_{\mathbf{y}=\hat{\mathbf{y}}} = 0$ or, equivalently, $f_i(t, \hat{y}_1, ..., \hat{y}_n) = 0$, for $i = 1, ..., n$. Notice that this description is valid even for systems which are more than first order in time since one could break a higher derivative equation in several first order equations. For instance, the system given by

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = f(t, y, y'), \tag{A.2}$$

can be rewritten as

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}t} &= Y \\ \frac{\mathrm{d}Y}{\mathrm{d}t} &= f(t, y, Y).\end{aligned} \tag{A.3}$$

Let us now consider a small perturbation around the equilibrium such that $y_i = \hat{y}_i + \epsilon_i$. We want to know if such small perturbation ϵ_i will grow or decay in time. If it decreases the system is stable under small perturbations, otherwise the system is deemed unstable. As ϵ_i are small perturbations, we can write a first order Taylor expansion of f_i around the equilibrium

$$f_i(t, \hat{y}_1 + \epsilon_1, ..., \hat{y}_n + \epsilon_n) \approx f_i(t, \hat{y}_1, ..., \hat{y}_n) + \sum_{j=1}^{n} \left.\frac{\mathrm{d}f_i}{\mathrm{d}y_j}\right|_{\mathbf{y}=\hat{\mathbf{y}}} \epsilon_j = \sum_{j=1}^{n} \left.\frac{\mathrm{d}f_i}{\mathrm{d}y_j}\right|_{\mathbf{y}=\hat{\mathbf{y}}} \epsilon_j \tag{A.4}$$

where the last equality comes from the fact that $f_i(t, \hat{y}_1, ..., \hat{y}_n) = 0$ at the equilibrium. In vector notation this can be rewritten as $f_i(t, \hat{y}_1{+}\epsilon_1, ..., \hat{y}_n{+}\epsilon_n) \approx \vec{\nabla} f_i\Big|_{\mathbf{y}=\hat{\mathbf{y}}} \cdot \vec{\epsilon}$. Putting all the n equation together the system takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\epsilon_1 \\ \vdots \\ \epsilon_n\end{pmatrix} = \left.\begin{pmatrix}\frac{\mathrm{d}f_1}{\mathrm{d}y_1} & \cdots & \frac{\mathrm{d}f_1}{\mathrm{d}y_n} \\ \vdots & \ddots & \vdots \\ \frac{\mathrm{d}f_n}{\mathrm{d}y_1} & \cdots & \frac{\mathrm{d}f_n}{\mathrm{d}y_n}\end{pmatrix}\right|_{\mathbf{y}=\hat{\mathbf{y}}} \begin{pmatrix}\epsilon_1 \\ \vdots \\ \epsilon_n\end{pmatrix} \quad \Leftrightarrow \quad \frac{\mathrm{d}\vec{\epsilon}}{\mathrm{d}t} = \mathcal{J} \cdot \vec{\epsilon}, \tag{A.5}$$

where $\mathcal{J}$ is the Jacobian matrix of the system evaluated at the equilibrium. With the help of a unitary transformation we can diagonalize the Jacobian matrix. Suppose $\mathcal{V}$ a unitary matrix such that $\mathcal{V}^{\dagger}\mathcal{J}\mathcal{V} = \mathcal{J}'$, where $\mathcal{J}'$ is the diagonalized Jacobian matrix. Let us define $\mathcal{V}^{\dagger}\,\vec{\epsilon} \equiv \vec{\xi}$. This way ξ_i is the projection of $\vec{\epsilon}$ in the i-th eigenvector of $\mathcal{J}$. Multiplying by $\mathcal{V}^{\dagger}$ and introducing $\mathcal{I} = \mathcal{V}^{\dagger}\mathcal{V}$ we can diagonalize the system, obtaining

$$\frac{\mathrm{d}\vec{\epsilon}}{\mathrm{d}t} = \mathcal{J} \cdot \vec{\epsilon} \quad \Rightarrow \quad \mathcal{V}^{\dagger}\frac{\mathrm{d}\vec{\epsilon}}{\mathrm{d}t} = \mathcal{V}^{\dagger}\mathcal{J}\mathcal{V}^{\dagger}\mathcal{V} \cdot \vec{\epsilon} \quad \Rightarrow \quad \frac{\mathrm{d}\vec{\xi}}{\mathrm{d}t} = \mathcal{J}' \cdot \vec{\xi} \tag{A.6}$$

This equation has the simple solution

$$\vec{\xi}(t) = \exp\left(\mathcal{J}'t\right) \cdot \vec{\xi}(0). \tag{A.7}$$

$\mathcal{J}'$ is a diagonal matrix, with the eigenvalues λ_i as elements. Therefore,

$$\exp\left(\mathcal{J}'t\right) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{pmatrix}. \tag{A.8}$$

So the projection of the perturbation in the direction of each eigenvector will grow independently of the other projections. We should make sure that no component of any arbitrary perturbation will grow. This can be done by requiring that all real parts of λ_i are smaller than 0, yielding the condition

$$\forall_i \ \operatorname{Re}(\lambda_i) < 0\,. \tag{A.9}$$

This way, any small perturbation will decrease exponentially in time and the system will return to its original state. If the imaginary part of eigenvalue is nonzero the system will oscillate while returning to the equilibrium.

Now let us see what happens in a corresponding time discrete system. A time discrete system can be generally written as

$$\begin{aligned} y_1(t+1) &= f_1\left(t, y_1(t), ..., y_n(t)\right) \\ &\vdots \\ y_n(t+1) &= f_n\left(t, y_1(t), ..., y_n(t)\right). \end{aligned} \tag{A.10}$$

We will consider an equilibrium solution $\hat{y}_i$, such that $y_i(t+1) = y_i(t) \equiv \hat{y}_i$, subject to a small perturbation $\epsilon_i(t)$. $y_i(t)$ can then be written as $y_i(t) \equiv \hat{y}_i + \epsilon_i(t)$. Replacing this in A.10 and expanding the f_i functions around the equilibrium one can find

$$\begin{aligned} \hat{y}_1 + \epsilon_i(t+1) &= \hat{y}_1 + \sum_{j=1}^{n} \left.\frac{\mathrm{d}f_1}{\mathrm{d}y_j}\right|_{\mathbf{y}=\hat{\mathbf{y}}} \epsilon_j(t) \\ &\vdots \\ \hat{y}_n + \epsilon_i(t+1) &= \hat{y}_n + \sum_{j=1}^{n} \left.\frac{\mathrm{d}f_1}{\mathrm{d}y_j}\right|_{\mathbf{y}=\hat{\mathbf{y}}} \epsilon_j(t). \end{aligned} \tag{A.11}$$

Therefore, similarly to before we can write

$$\vec{\epsilon}(t+1) = \mathcal{J} \cdot \vec{\epsilon}(t) \quad \Rightarrow \quad \vec{\xi}(t+1) = \mathcal{J}' \cdot \vec{\xi}(t), \tag{A.12}$$

where the expression on the right corresponds to diagonalized version of the system. This system's solution is

$$\vec{\xi}(t) = (\mathcal{J}')^t \cdot \vec{\xi}(0). \tag{A.13}$$

The power of matrix a diagonal matrix is simply the matrix with the powers of the diagonal entries. Thus, one has

$$(\mathcal{J}')^t = \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^t \end{pmatrix}. \tag{A.14}$$

As we want to guarantee that the perturbation is damped in time, we should require that the absolute value of the eigenvalues is smaller than one. So the relevant condition can be expressed as

$$\forall_i \; |\lambda_i| < 1\,. \tag{A.15}$$

APPENDIX B – MORAN PROCESS

Very large populations can be accounted for by differential equations. Consider, for simplicity, the growth of a haploid population[1]. If the population has growing rate a the evolution of the population size can be described by the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a\,x. \tag{B.1}$$

Let us now introduce selection in the model. If we have two competing strains with growing rates a and b the fraction of individuals x_a and x_b of each strain evolves according to the system

$$\frac{\mathrm{d}x_a}{\mathrm{d}t} = x_a(a - \phi), \tag{B.2}$$

$$\frac{\mathrm{d}x_b}{\mathrm{d}t} = x_b(b - \phi). \tag{B.3}$$

where ϕ is introduced to guarantee that $x_a + x_b = 1$. Simple algebra leads us to the conclusion that $\phi = ax_a + bx_b$, i.e., ϕ is the average fitness of the population. Since we have the condition $x_a + x_b = 1$, the system can be reduced to a single equation

$$\frac{\mathrm{d}x_a}{\mathrm{d}t} = x_a(1 - x_a)(a - b), \tag{B.4}$$

which has the solution

$$x_a(t) = \frac{x_{a,0}\, e^{(a-b)t}}{1 - x_{a,0}(1 - e^{(a-b)t})} \tag{B.5}$$

[1] A haploid organism is an organism that has only one copy of each chromosome per cell.

where $x_{a,0}$ is the initial population fraction of strain a. From this expression we can see that, given that the initial population does not consist of a single strain, i.e. $x_{a,0} \neq 0$ and $x_{a,0} \neq 1$, the evolution of the population depends only on the growth rates a and b. Namely, for large t, if $a > b$ the strain a will always dominate and if $a < b$ the population is dominated by the strain b. The ultimate fate of the population is found deterministically independently of the initial proportions of the strains.

This approach reveals quite unsatisfactory to deal with finite populations since it dismisses all stochastic effects and, therefore, it does not capture the population drift effect. Stochastic effects are an unavoidable ingredient of any finite population due to the random sampling of the population for reproduction. This shortcoming can be addressed in various ways. Historically, two simple models have been widely employed to deal with this limitation: the Wright-Fisher process and the Moran process. Both models consider a constant and finite population with size N. In the Wright-Fisher model, the population is replaced every time step by a new population sampled from the parent population. The reproduction probability is proportional to the parent's fitness. The Moran process was proposed by Patrick Moran in the article *Random processes in genetics* in 1958 [135]. It aimed to change the Wright-Fisher model by introducing overlapping generations. The reproduction process is dealt with differently in Moran process. At every time step one individual is randomly chosen to reproduce and replaces one individual of the original population. This way one step of Wright-Fisher corresponds to N steps in Moran process. Also, in Moran process the number of individuals of a certain strain cannot change by more than one in a given time step.

Recovering the example of two strains with constant fitness a and b, let us say strain a has i individuals and, consequently, there are $N - i$ individuals of strain b. There are three distinct possibilities: i is kept if the individual that reproduces and the eone that dies are from the same strain, i can increase by one if a strain a individual reproduces and one from strain b dies and i can decrease by one if an individual from strain b reproduces and one from strain a dies. As the reproduction probability is proportional to the fitness of each individual the probability of one individual of strain a to reproduce is given by

$$P_{r,a} = \frac{ia}{ia + (N-i)b}, \tag{B.6}$$

and the probability that one individual b reproduces is the complementar, i.e.,

$$P_{r,b} = \frac{(N-i)b}{ia + (N-i)b}, \tag{B.7}$$

such that $P_{r,a} + P_{r,b} = 1$. For the death probabilities, we have

$$P_{d,a} = \frac{i}{N}, \tag{B.8}$$

$$P_{d,b} = \frac{N-i}{N}. \tag{B.9}$$

Therefore, the probability of the transition $i \to i+1$, $P_{i \to i+1}$, is given by

$$P_{i \to i+1} = P_{r,a} P_{d,b} = \frac{ia}{ia + (N-i)b} \frac{N-i}{N}, \tag{B.10}$$

and the remaining two possibilities are

$$P_{i \to i} = P_{r,a} P_{d,a} + P_{r,b} P_{d,b} = \frac{ia}{ia + (N-i)b} \frac{i}{N} + \frac{(N-i)b}{ia + (N-i)b} \frac{N-i}{N}, \tag{B.11}$$

and

$$P_{i \to i-1} = P_{r,b} P_{d,a} = \frac{(N-i)b}{ia + (N-i)b} \frac{i}{N}. \tag{B.12}$$

As expected, $P_{i \to i-1} + P_{i \to i} + P_{i \to i+1} = 1$, therefore $P_{i \to i} = 1 - P_{i \to i-1} - P_{i \to i+1}$. The system has two absorbing states[2], that correspond to $i = 0$ and $i = N$. What is the probability of reaching the state $i = N$, i.e., a dominates the population, if we start from a state with a single a individual? This probability, called fixation probability, is important to study the evolution of population where a mutant appears since it tells us whether it is probable that this new mutant will be lost or dominate the population. Notice that, in the situation described by the differential equation, this probability was 1 if $a > b$ and 0 if $a < b$. In a finite population there will be drift effects that make the fixation probability nonzero even when the strain is

[2] An absorbing state is a state that once entered cannot be left. In our example the states $i = 0$ and $i = N$ are absorbing states since they correspond to the extinction of strain a and b, respectively.

disadvantageous. Let us introduce the notation x_i for the probability of reaching the state N starting from i. Naturally, we have

$$\begin{aligned} x_0 &= 0, \\ x_i &= P_{i-1\to i}\, x_{i-1} + P_{i\to i}\, x_i + P_{i+1\to i}\, x_{i+1}, \\ x_N &= 1. \end{aligned} \tag{B.13}$$

Let us now introduce a new set of variables

$$y_i = x_i - x_{i-1}, \tag{B.14}$$

and define an auxiliary quantity $R \equiv \frac{P_{i-1\to i}}{P_{i+1\to i}} = \frac{a}{b}$. It is easy to realize that

$$\sum_{i=1}^{N} y_i = x_1 - x_0 + x_2 - x_1 + \cdots + x_N - x_{N-1} = x_N - x_0 = 1. \tag{B.15}$$

Using the equation B.13 one can find

$$\begin{aligned} x_{i+1} &= \frac{-P_{i\to i}x_i - P_{i-1\to i}x_{i-1} + x_i}{P_{i+1\to i}} = \frac{\left(P_{i-1\to i} + P_{i+1\to i} - 1\right) x_i - P_{i-1\to i}x_{i-1} + x_i}{P_{i+1\to i}} \\ &= x_i + \frac{P_{i-1\to i}}{P_{i+1\to i}} \left(x_i - x_{i-1}\right) = x_i + R\left(x_i - x_{i-1}\right). \end{aligned} \tag{B.16}$$

Applying this we obtain

$$y_{i+1} = x_{i+1} - x_i = x_i + R\left(x_i - x_{i-1}\right) - x_i = Ry_i. \tag{B.17}$$

Since $y_1 = x_1 - x_0 = x_1$, it follows that $y_k = R^{k-1}x_1$. Therefore,

$$\sum_{i=1}^{N} y_i = 1 \Rightarrow x_1 \sum_{i=1}^{N} R^{i-1} = 1 \Rightarrow x_1 = \frac{1}{\sum_{i=1}^{N} R^{i-1}} = \frac{1-R}{1-R^N}. \tag{B.18}$$

This way, the fixation probability of a mutant with fitness a in a population of individuals with fitness b is

$$P_{fix} = \frac{1-R}{1-R^N}, \tag{B.19}$$

where R is the relative fitness $R = a/b$ and N is the population size. We can see that now the fixation probability depends not only on the fitnesses a and b, but also on the population size N. Even a disadventageous mutant has a nonzero fixation probability.

This appendix only attempts to provide a very simple introduction to the Moran process and the concept of drift. It does not deal with many aspects that are highly relevant in the dynamics of finite populations, such as frequency dependent selection, mutation, migration and population structure. For a somewhat deeper introduction check, for example, the book *Evolutionary dynamics*, by Martin A. Nowak [58].

APPENDIX C – GILLESPIE ALGORITHM

Gillespie algorithm is a Monte-Carlo algorithm first introduced mathematically by Joseph Doob in 1942 [136, 137]. It was later reintroduced and popularized by David Gillespie in the context of chemical reactions simulations in 1976 [129, 138]. This algorithm allows exact simulations of chemical reactions with arbitrary number molecules, including low amounts of molecules, that would not be correctly described by continuum differential equations.

Gillespie algorithm proceeds in two steps: first calculates the time to the next reaction based on the total reaction rate, or propensity, of the system; then chooses each reaction has happened based on the relative reaction rate of each reaction, and returns to the previous step. This provides a simple algorithm capable of accurately describing the time evolution of chemical systems. Each run of the algorithm provides a possible trajectory of the system in time and performing multiple runs allows one to know the statistical properties of the system.

Although introduced in the context of chemical reactions, this algorithm can be used in a variety of contexts, whenever the state of the system can be described by rates of transition between states that remain constant between changes.

The algorithm assumes that in a given time interval t to $t + dt$ each possible process μ has an average probability $k_\mu \mathrm{d}t$ to happen. Let us call $P(\tau, \mu)\mathrm{d}\tau$ the probability that the next process is μ and happens in the time interval between $t + \tau$ and $t + \tau + \mathrm{d}\tau$. Then, $P(\tau, \mu)\mathrm{d}\tau$ is given by $P(\tau, \mu)\mathrm{d}\tau = P_0(\tau)k_\mu \mathrm{d}\tau$, i.e., the probability that no process occurs during the interval τ times the probability of the process μ occurs in the interval $t + \tau$ to $t + \tau + \mathrm{d}\tau$. One can then divide the interval τ in n smaller equal time intervals of duration $\epsilon = \tau/n$. As the probability of no event happening in an infinitesimal time interval $\mathrm{d}t$ is $1 - \sum_\nu k_\nu \mathrm{d}t$ (1 minus the probability of any event occuring) we can write $P_0(\tau) = \left(1 - \sum_\nu k_\nu \frac{\tau}{n} + \mathcal{O}(\tau/n)\right)^n$. Taking the limit of infinite subintervals we obtain $P_0(\tau) = \lim_{n\to\infty} \left(1 - \sum_\nu k_\nu \frac{\tau}{n}\right)^n = e^{-\sum_\nu k_\nu \tau}$. Therefore the probability of the process μ occuring in the given conditions is $P(\tau, \mu)\mathrm{d}\tau = e^{-\sum_\nu k_\nu \tau} k_\mu \mathrm{d}\tau$.

APPENDIX D – MAXIMUM VIABILITY FOR ONE TRADEOFF

Knowing the tradeoff matrix we can estimate the viability under some assumptions. Let us first analyze the situation with one tradeoff and no specialization. If there is no specialization all the regulatory genes y_{ik} will take null values. This eliminates the costs of regulation but subjects the organism to tradeoff costs. In order to achieve a nonzero fitness the organism will have to decrease the activity of some of its main genes Y_i. With this condition Eq. 3.5 reduces to the form

$$\varphi_{ik} = \prod_j |1 - \delta_{ij} - Y_j|^{T_{ij}}. \tag{D.1}$$

Note that now that the cells are in an undifferentiated state, φ_{ik} does not depend on k anymore since all the cells behave the same independently of the stimuli they are subject to. Thus, in this calculation we drop the index k and take directly $\varphi_i = \varphi_{ik}$. Assuming only one tradeoff the tradeoff matrix will have only one nonzero entry outside the diagonal. Let us assume, for simplicity, that the only nonzero off-diagonal entry of the tradeoff matrix is T_{12}. This does not entail loss of generality since we are ascribing the same importance to each function, making the model symmetric with respect to the position of the nonzero entry in the tradeoff matrix. So the tradeoff matrix we are considering has the following structure

$$\mathbf{T} = \begin{bmatrix} s & s^* & 0 & \cdots & 0 \\ 0 & s & 0 & \cdots & 0 \\ 0 & 0 & s & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s \end{bmatrix}. \tag{D.2}$$

Therefore we will have the following φ_{ik} values

$$\begin{aligned} \varphi_1 &= Y_1^s \, (1-Y_2)^0 \ldots (1-Y_n)^0 = Y_1^s, \\ \varphi_2 &= (1-Y_1)^{s^*} \, Y_2^s \, (1-Y_3)^0 \ldots (1-Y_n)^0 = (1-Y_1)^{s^*} \, Y_2^s, \\ \varphi_{i>2} &= (1-Y_1)^0 \ldots (1-Y_{i-1})^0 \, Y_i^s \, (1-Y_{i+1})^0 \ldots (1-Y_n)^0 = Y_i^s. \end{aligned}$$

A factor $(1-Y_1)^{s^*}$ appears, suppressing the output of the task 2 when performed concomitantly with the task 1. The viability will be

$$v = \left(\prod_i \varphi_i \right)^{1/n} = \left[Y_1^s \, (1-Y_1)^{s^*} \, Y_2^s \, \prod_{i>2} Y_i^s \right]^{1/n} \tag{D.3}$$

An evolutionary system will search for the maxima of the fitness. As v is a strictly increasing function of the Y_i for $i > 1$ it presents no constraint to fitness maximization and take their maximum possible value $Y_{i>1} = 1$. The same does not happen with Y_1 due to the tradeoff restrictions. We should maximize the viability with respect to Y_1. This corresponds to finding the zero of the derivative

$$\begin{aligned} &\left. \frac{\mathrm{d}v}{\mathrm{d}Y_1} \right|_{Y_1=\overline{Y}} = 0 \\ \Rightarrow &\left[\overline{Y}^s \, (1-\overline{Y})^{s^*} \right]^{1/n-1} \left[s\overline{Y}^{s/n-1}(1-\overline{Y})^{s^*/n} - s^*\overline{Y}^{s/n}(1-\overline{Y})^{s^*/n-1} \right] = 0. \end{aligned} \tag{D.4}$$

Taking into account that $\overline{Y}$ cannot be 0 or 1, owing to the fact that the viability would become 0, we can simplify this expression to

$$s(1-\overline{Y}) - s^*\overline{Y} = 0 \quad \Rightarrow \quad \overline{Y} = \frac{s}{s+s^*}. \tag{D.5}$$

This expression can now be replaced in the viability to find the maximum viability achievable without specilization

$$\overline{v}=\left[\left(\frac{s}{s+s^{*}}\right)^{s}\left(1-\frac{s}{s+s^{*}}\right)^{s^{*}}\right]^{1/n}=\left[\frac{s^{s}s^{*s^{*}}}{(s+s^{*})^{s+s^{*}}}\right]^{1/n}. \tag{D.6}$$

It is interesting to compare the maximum viability without specialization to the viability obtained with specialization. The full specialization regime should arise from natural selection in the case of very strong tradeoffs. When the specilization is in place the tradeoffs no longer reduce the value of Y_i. Therefore, we can assume that it is in the maximum of the viability $Y_i = 1$, for all i. With this condition, the Eq. 3.5 simplifies to

$$\varphi_{ik}=\prod_{j}|y_{jk}-\delta_{ij}|^{T_{ij}}\,c(y_{jk}). \tag{D.7}$$

We will consider the same tradeoff matrix chosen before, described in Eq. D.2. All y_{ik}'s will be considered zero, unless that leads to a zero fitness, in which case we will consider them one. This leads to total specialization in the functions that present tradeoffs. The cells remain generalists regarding the tasks that are not affected by tradeoffs. In the case of the tradeoff matrix under consideration this means that $y_{12} = 1$ and all the remaining regulatory genes are null. After some analysis one can find

$$\begin{aligned}
\varphi_{11} &= |y_{11}-1|^{s}c(y_{11})\ (y_{21})^{0}c(y_{21})\ldots(y_{n1})^{0}c(y_{n1})=1,\\
\varphi_{12} &= |y_{12}-1|^{s}c(y_{12})\ (y_{22})^{0}c(y_{22})\ldots(y_{n2})^{0}c(y_{n2})=0,\\
\varphi_{21} &= (y_{11})^{s^{*}}c(y_{11})\ |y_{21}-1|^{s}c(y_{21})\ldots(y_{n1})^{0}c(y_{n1})=0,\\
\varphi_{22} &= (y_{12})^{s^{*}}c(y_{12})\ |y_{22}-1|^{s}c(y_{22})\ldots(y_{n2})^{0}c(y_{n2})=c(1),\\
\varphi_{ii>2} &= (y_{1i})^{0}c(y_{1i})\ldots|y_{ii}-1|^{s}c(y_{ii})\ldots(y_{n2})^{0}c(y_{n2})=1,\\
\varphi_{1k>2} &= |y_{1k}-1|^{s}c(y_{1k})\ldots(y_{nk})^{0}c(y_{nk})=1,\\
\varphi_{(i>2)\,1} &= (y_{11})^{0}c(y_{11})\ldots(y_{n1})^{0}c(y_{n1})=1,\\
\varphi_{2k>2} &= (y_{1k})^{s^{*}}c(y_{11})|y_{2k}-1|^{s}c(y_{2k})\ldots(y_{nk})^{0}c(y_{nk})=0,\\
\varphi_{(i>2)\,2} &= (y_{12})^{0}c(y_{12})\ldots(y_{n2})^{0}c(y_{n2})=c(1).
\end{aligned} \tag{D.8}$$

In the calculation of these expressions the condition $c(0) = 1$ was used. As we can see, while in the case of generalists the result did not depend on the stimuli, in the limit of fully specialized cells the result becomes independent of the tradeoff strength. The values of φ_i are given by

$$\begin{aligned}
\varphi_1 &= \frac{1}{n}\sum_k \varphi_{1k} = \frac{1}{n}\left[1 + 0 + (n-2)\times 1\right] = \frac{1}{n}\left(n-1\right),\\
\varphi_2 &= \frac{1}{n}\sum_k \varphi_{2k} = \frac{1}{n}\left[0 + c(1) + (n-1)\times 0\right] = \frac{1}{n}c(1),\\
\varphi_{i>2} &= \frac{1}{n}\sum_k \varphi_{ik} = \frac{1}{n}\left[c(1) + (n-1)\times 1\right] = \frac{1}{n}\left[n-1+c(1)\right].
\end{aligned}$$

Therefore, the maximum viability of an organism with total specialization is

$$\begin{aligned}
\overline{v} = \left(\prod_i \varphi_i\right)^{1/n} &= \left[\frac{1}{n}\left(n-1\right)\frac{1}{n}c(1)\prod_i \frac{1}{n}\left[n-1+c(1)\right]\right]^{1/n}\\
&= \frac{1}{n}\left[\left(n-1\right)c(1)\left[n-1+c(1)\right]^{n-2}\right]^{1/n}.
\end{aligned} \tag{D.9}$$

The general expression, which does not assume neither pure generalist nor total specialization, can be found following the same method. After some calculations, one obtains

$$\begin{aligned}
\overline{v} =& \frac{1}{n}\left\{Y_1^s\left[n-1+\left(1-y_{12}\right)^s c(y_{12})\right]\right.\\
&\times\left[\left(n-1\right)\left(1-Y_1\right)^{s^*} + \left(1-(1-y_{12})Y_1\right)^{s^*} c(y_{12})\right]\\
&\times\left.\left[n-1+c(y_{12})\right]^{n-2}\right\}^{1/n}.
\end{aligned} \tag{D.10}$$

This expression includes the possibility of partial specialization. It is an interesting exercise to verify that the previously discussed expressions are recovered from the general one when the appropriate limits are considered.