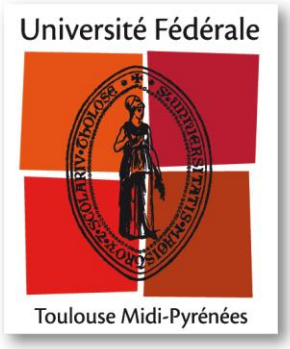


THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par

Sébastien Maignan

Le 30 août 2018

Study of conditions for the emergence of cellular communication using self-adaptive multi-agent systems

---

École doctorale et discipline ou spécialité

ED MITT : Domaine STIC : Intelligence Artificielle

Unité de recherche

Institut de Recherche en Informatique de Toulouse

Directrice(s) ou Directeur(s) de Thèse

Pierre Glize, Carole Bernon

Jury

Marie BEURTON-AIMAR	Maître de Conférences, HdR, Université de Bordeaux	Rapporteur
Vincent RODIN	Professeur des Universités, Université de Bretagne Occidentale	Rapporteur
Arnaud DUCRUIX	Professeur Emérite, Université Paris Descartes	Examinateur
Bernard FRANCÈS	Professeur Emérite, Université de Toulouse	Examinateur
Marie-Pierre GLEIZES	Professeur des Universités, Université de Toulouse	Examinatrice
Guillaume HUTZLER	Maître de Conférences, HdR, Université d'Evry	Examinateur
Carole BERNON	Maître de Conférences, Université de Toulouse	Co-Directrice
Pierre GLIZE	Ingénieur d'Etude, HdR, CNRS	Directeur
Jean-Pierre MANO	Data Scientist, Brennus Analytics	Invité

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par

Sébastien Maignan

Le 30 août 2018

Study of conditions for the emergence of cellular communication using self-adaptive multi-agent systems

---

École doctorale et discipline ou spécialité
ED MITT : Domaine STIC : Intelligence Artificielle

Unité de recherche
Institut de Recherche en Informatique de Toulouse

Directrice(s) ou Directeur(s) de Thèse
Pierre Glize, Carole Bernon

Jury

Marie BEURTON-AIMAR	Maître de Conférences, HdR, Université de Bordeaux	Rapporteur
Vincent RODIN	Professeur des Universités, Université de Bretagne Occidentale	Rapporteur
Arnaud DUCRUIX	Professeur Emérite, Université Paris Descartes	Examinateur
Bernard FRANCÈS	Professeur Emérite, Université de Toulouse	Examinateur
Marie-Pierre GLEIZES	Professeur des Universités, Université de Toulouse	Examinatrice
Guillaume HUTZLER	Maître de Conférences, HdR, Université d'Evry	Examinateur
Carole BERNON	Maître de Conférences, Université de Toulouse	Co-Directrice
Pierre GLIZE	Ingénieur d'Etude, HdR, CNRS	Directeur
Jean-Pierre MANO	Data Scientist, Brennus Analytics	Invité

Remerciements

Tout d'abord je souhaiterai remercier les membres du jury Marie Beurton-Aimar, Vincent Rodin, Arnaud Ducruix, Bernard Francès, Guillaume Hutzler et Jean-Pierre Mano d'avoir accepté d'évaluer ce travail.

Je voudrais remercier Marie-Pierre Gleizes de m'avoir accueilli chez SMAC bien que mon sujet sortait un peu de la thématique du laboratoire ainsi que de sa gentillesse.

Je tiens à remercier Pierre Glize et Carole Bernon qui ont accepté de me guider pendant cette thèse. Je me souviendrais toujours de nos sessions de travail où nous essayions d'imaginer les « motivations » des cellules primitives.

J'aimerais aussi remercier mon épouse Marie-José et mes enfants Julien, Karine et Alexandre de leur soutien inconditionnel.

Enfin je voudrais exprimer ma gratitude à tous les membres de l'équipe SMAC. Ces quelques années passées en votre compagnie ont été une expérience inoubliable à bien des égards. Que ce soit pour les discussions scientifiques, philosophiques ou pour la bonne humeur permanente, j'en ai apprécié chaque seconde. Cependant comme je ne suis pas certain de trouver les mots justes pour vous tous, j'ai trouvé une boite usagée mais magique qui contient le remerciement parfait pour chacun d'entre vous. Elle possède des trous d'aérations car à une époque il est dit qu'elle a contenu un mouton… ☺.

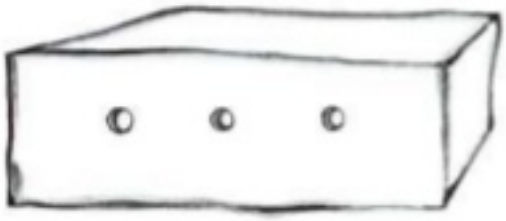


Table of contents

CHAPTER 1. INTRODUCTION

The english version starts page 5.

42! Ceci est censé être la réponse ultime à la vie, à l'univers et au reste (Adams 1980). Cela pourrait bien être le cas, mais il y a encore de nombreux domaines où cela ne nous aide pas dans nos recherches quotidiennes. 42 semble émerger de nulle part et en fait, l'émergence est un véritable phénomène dans notre univers. Dans la vie en particulier, nous pouvons observer les effets de l'émergence partout où nous portons notre attention. De l'assemblage d'atomes pour former de petites molécules, aux protéines et à l'ADN. Des cellules égoïstes aux communautés cellulaires des formes de vie « supérieures ». Des communautés de neurones à la conscience. Et des communautés de consciences aux sociétés. Où s'arrête-t-elle ? Nous ne le savons pas encore. Comment cela fonctionne-t-il ? Nous essayons toujours de le comprendre, mais plusieurs axes de recherche progressent bien.

Un exemple d'émergence dans la nature est d'un intérêt particulier dans le cadre de ce travail. Il se produit lors de la transition d'une collection de cellules individuelles à un assemblage de cellules hautement spécialisées et interdépendantes travaillant ensemble, c'est-à-dire des organismes multicellulaires. Plusieurs questions sont soulevées par ce changement de comportement. Pourquoi cela se produit-il ? Les cellules égoïstes individuelles sont encore présentes de nos jours et prospèrent. Même dans un corps humain normal, il y a plus de cellules non humaines (cellules égoïstes individuelles) que de cellules humaines. De plus, les organismes multicellulaires ont besoin de coordination cellulaire, d'organisation et d'un niveau plus élevé de complexité de leurs processus internes. Cela représente un coût élevé pour les cellules et est beaucoup plus sensible aux défaillances que les organismes simples. Alors pourquoi s'embêter ? Comme c'est souvent le cas avec la nature (et la loi de Murphy), si cela peut arriver, cela arrivera. Et comme l'inconvénient d'être un organisme multicellulaire n'est pas toujours un défaut critique, nous pouvons encore rencontrer des représentants de cette forme de vie sur Terre. Néanmoins, ils ne représentent qu'une part négligeable du nombre total d'organismes et d'espèces et leur avenir pourrait bien être incertain puisqu'ils sont si peu nombreux et fragiles.

Une autre question est de savoir comment cela se produit-il ? C'est une question très délicate et il n'y a pas de consensus général sur le processus. Ce qui est sûr, c'est qu'il est apparu et a disparu plusieurs fois au cours de l'histoire de la Terre. Donc, encore une fois, ce changement ne représente pas la flèche « naturelle » de l'évolution, mais c'est une option possible et viable selon les circonstances. En d'autres termes, une complexité accrue n'est pas la clé du succès. L'adaptation à l'environnement semble être la seule vérité sur le fonctionnement de l'évolution. De cette façon et sous certaines conditions, les organismes multicellulaires sont mieux adaptés à l'environnement. En particulier, la capacité de récolter de la nourriture dans de plus grandes zones et la protection contre les prédateurs semblent être des atouts importants de ce type d'organismes.

En résumé, les organismes multicellulaires sont apparus parce qu'il s'agissait d'une solution viable dans des environnements spécifiques et peuvent même, dans certains cas, présenter des avantages concurrentiels pour la survie.

Quelles sont les caractéristiques minimales communes de ces organismes multicellulaires ? Les cellules sont interdépendantes et souvent spécialisées. Un autre aspect apparemment très important est la parenté, c'est-à-dire un degré élevé de similitude entre les membres de la communauté cellulaire et le fait qu'ils travaillent tous de la même manière. Ils sont également capables de se coordonner pour atteindre un objectif global comme se diriger vers une proie. Néanmoins, aucune cellule individuelle ne connaît cet objectif global ; le mouvement est une propriété émergente du collectif cellulaire.

Pour atteindre les objectifs communautaires, les cellules doivent communiquer entre elles d'une manière ou d'une autre. Ce que l'on observe dans les organismes supérieurs modernes est un système élaboré de molécules spécialisées utilisées comme messagers entre les cellules. À la surface de chaque cellule, des récepteurs spécifiques sont présents pour détecter et transférer l'information à travers la membrane externe de la cellule. Chaque type de cellule ne présente qu'un sous-ensemble de récepteurs possibles afin de filtrer les messages d'intérêt.

Une partie importante de la biologie cellulaire et de la biochimie consiste à comprendre la nature de ces messages, à les cataloguer et à analyser leur rôle. Il existe plus de 200 molécules de ce type classées dans différentes familles. Cela présente un intérêt particulier pour l'industrie pharmaceutique puisque plus de la moitié de la pharmacopée actuelle dégrade ou améliore le transfert d'information entre les cellules. Ceci démontre l'importance des protocoles de communication dans la survie des organismes multicellulaires. Il est aussi assez surprenant de constater qu'il est souvent difficile d'attribuer une fonction spécifique à une molécule de signalisation donnée.

Les études in vivo et in vitro des molécules de signalisation peuvent être très difficiles et les résultats difficiles à interpréter. En particulier, il est tout à fait clair qu'une molécule donnée peut induire une réponse cellulaire dans des conditions expérimentales spécifiques et une réponse cellulaire différente pour d'autres conditions (voir https://www.nextprot.org pour des exemples). De manière plus compréhensible, un même messager semble aussi être capable d'induire des effets différents sur différents types de cellules.

Contribution de cette thèse

Une question intéressante est de savoir si le dispositif expérimental est responsable des différents comportements observés pour les mêmes messagers ou si ces divergences ont une signification plus profonde. Par exemple, il y a de plus en plus de preuves montrant que la nature de la réponse induite à une molécule dépend des contextes interne et externe de la cellule qui la reçoit. De plus, nous savons déjà que la nature optimise l'utilisation des ressources et qu'elle aurait pu faire la même chose avec les systèmes de communication cellulaire. Au lieu d'avoir une molécule messagère pour chaque action qu'une cellule peut effectuer, l'évolution pourrait avoir choisi un système où les combinaisons de molécules messagères transportent l'information pertinente pour la cellule et non les molécules elles-mêmes. Chaque messager devient un mot dans une phrase, mais seule la phrase complète a un sens. Cela permettrait de réutiliser les mêmes molécules dans des contextes différents. Est-ce la raison des étranges résultats expérimentaux qui sont parfois publiés ?

Afin de tester cette hypothèse de communication structurée, la simulation par ordinateur est bien adaptée puisqu'elle permet une maîtrise totale de l'environnement du système et devrait être beaucoup plus facile à mettre en place que les expériences in vitro. Simuler des colonies de cellules entières est encore chose lointaine, mais en utilisant des hypothèses appropriées pour simplifier le modèle, on peut obtenir des résultats très intéressants à partir d'expériences in silico. Cependant, lors de ces expériences, il est primordial d'utiliser des méthodes et des algorithmes qui n'introduisent aucun biais vers le phénomène émergent que l'on souhaite étudier. Toute méthode qui utilise une

fonction de coût pour évaluer la progression du système vers l'objectif prévu est donc interdite. Cela inclut les réseaux de neurones, les algorithmes génétiques et beaucoup d'autres algorithmes populaires. Sans introduire aucun biais, il est aussi indispensable d'explorer l'espace de paramètres du système de la manière la plus efficace possible puisque 4 milliards d'années ne sont pas disponibles pour expérimenter. Nous verrons que les systèmes multi-agents adaptatifs (AMAS) sont un cadre intéressant qui répond à ces deux attentes. Fondamentalement, l'approche AMAS utilise la coopération entre agents pour échantillonner de manière efficace l'espace des paramètres, et son mécanisme de décision locale évite toute possibilité de biais vers des objectifs prédéfinis à l'échelle du système. Il est également intéressant de noter que la coopération dans le domaine de la biologie est un trait essentiel qui a rendu possible la transition d'organismes unicellulaires à des organismes multicellulaires « supérieurs ».

Du côté de l'approche AMAS, ce travail explorera la notion de coopération dans un environnement où les agents ne disposent pas initialement de protocoles de communication pour coordonner leurs actions. Il est intéressant d'étudier si les principes de coopération et de similitude des agents sont suffisants pour faire émerger du système un « langage » de communication viable. L'auto-organisation et l'adaptation de la structure AMAS sont testées en dehors de leurs limites habituelles, ce qui pourrait étendre le champ d'application des AMAS.

Cette thèse décrit le développement d'un modèle simple mais suffisamment représentatif d'une collection de cellules. En utilisant la coopération et la parenté, ce modèle est utilisé pour explorer le potentiel d'homéostasie des ressources, le partage des ressources critiques et enfin l'émergence d'un protocole de communication de coordination. Parallèlement, et comme la puissance de calcul et le temps disponible sont limités, la coopération dans le cadre de l'approche AMAS est utilisée pour accélérer l'exploration spatiale des paramètres sans introduire de biais.

Organisation du manuscrit

Le manuscrit est organisé comme suit :

Chapitre 2 : Ce chapitre présente les contextes biologique et informatique de ce travail et les notions clés nécessaires à la construction du modèle de simulation cellulaire. Les mécanismes de communication observés dans des organismes multicellulaires réels sont décrits, puis l'état de l'art en biologie et en informatique est présenté. L'état de l'art consacré à la partie biologique discute d'expériences publiées soutenant l'hypothèse de cette thèse. Pour la partie informatique, les logiciels et algorithmes traitant des simulations cellulaires et multicellulaires sont présentés.

Chapitre 3 : Dans ce chapitre, CoCell, la contribution de ce travail, est présentée. Les choix effectués pour construire cette simulation sont justifiés avant d'établir les bases de sa mise en œuvre selon la théorie AMAS. Les composants de la simulation sont agentifiés et les rôles et comportements de ces agents sont introduits.

Chapitre 4 : CoCell1, la première mise en œuvre de la contribution est présentée dans ce chapitre. Sa performance à maintenir en vie un système multicellulaire, ainsi que sa stabilité et sa robustesse sont évaluées par différentes expériences.

Chapitre 5 : Ce chapitre enrichit CoCell1 avec des mutations cellulaires pour mettre en œuvre la deuxième partie de la contribution, CoCell2. L'impact de ces mutations sur la stabilité et l'adaptabilité du système est évalué.

Chapitre 6 : Enfin, la dernière instance, CoCell3, est décrite dans ce chapitre. Cette dernière étape permet d'étudier les conditions nécessaires et suffisantes pour observer l'émergence de la communication entre cellules à partir de comportements simples d'agents.

Chapitre 7 : Ce dernier chapitre conclut et présente les perspectives. Les implications des résultats de ce travail sur les organismes multicellulaires réels sont discutées. Enfin, des orientations possibles pour des travaux futurs dans le domaine de la simulation multicellulaire sont proposées.

42! This is supposed to be the ultimate answer to life, the universe and everything (Adams 1980). It might well actually be the case, but there are still many areas where this does not help us in our daily research. 42 seemingly emerges from nowhere and actually emergence is a true phenomenon in our universe. In life in particular we can observe its effects wherever we focus our attention. From the assembly of atoms to form small molecules, to proteins and DNA. From egoist cells to cell communities of "higher" life forms. From neuron communities to consciousness. And from consciousness communities to societies. Where does it stop? We do not know yet. How does it work? We are still trying to understand, but several lines of research are making good progress.

One occurrence of emergence in nature is of special interest in this work. It happens during the transition from a collection of individual cells to an assembly of highly specialized interdependent cells working together *i.e.* multicellular organisms. Several questions are raised by this behavioral change. Why does it happen? Individual egoist cells are still present nowadays and thrive. Even in a normal human body there are more non-human cells (egoist single cells) than human cells. Furthermore, multicellular organisms require cell coordination, organization and a higher level of complexity of their inner processes. This represents a high cost for cells and is much more susceptible to failure than simple organisms. So why bother? As it is often the case with Nature (and Murphy's law) if it can happen it will happen. And since drawback of being a multicellular organism is not always a critical flaw we can still encounter some representatives of this form of life on Earth. Nonetheless, they represent only a negligible part of the total number of organisms and species and their future might well be uncertain since they are so few and fragile.

Another question is how does it happen? This is a very tricky question and there is no general consensus on the process. One sure thing is that it appeared and disappeared several times during Earth history. So again this change does not represent the "natural" arrow of evolution but it happens to be a possible and viable option depending on the circumstances. In other words, increased complexity is not key to success. Adaptation to the environment seems to be the only truth about the way evolution works. In this way and in certain conditions multicellular organizations are better adapted to the environment. In particular, the ability to gather food in larger zones and protection from predators appear to be important assets of this type of organisms.

To summarize, multicellular organisms appeared because it was a viable solution in specific environments and may even present some survival competitive advantages in some cases.

Now, what are the common minimal characteristics of these multicellular organisms? Cells are interdependent and often specialized. Also a seemingly very important aspect is kinship *i.e.* high degree of relatedness between members of the cell community and the fact that they all work in the same way. They are also able to coordinate themselves to perform a global objective like moving towards a prey. Nevertheless, each individual cell is unaware of this global goal. The motion is an emergent property of the cell collective.

In order to achieve community level goals, cells need to communicate together in one way or another. What is observed in modern day higher organisms is an elaborate system of specialized molecules used as messengers between cells. On the surface of each cell, messenger specific receptors are present to detect and transfer the information through the cell outer membrane. Each cell type only presents a subset of possible receptors in order to filter the messages of interest.

An important part of cell biology and biochemistry is understanding the nature of these messages, cataloging them and understanding their role. There exist more than 200 such molecules classified in different families. This is of particular interest for the pharmaceutical industry since more than half of current pharmacopoeia either antagonize or enhance the transfer of information

between cells. This demonstrates the importance of the communication protocols in the survival of multicellular organisms. Surprisingly enough, it is often difficult to attribute a specific function to a given signaling molecule.

In vivo and *in vitro* studies of signaling molecules can be quite challenging and the results difficult to interpret. In particular, it is quite clear that a given molecule can induce a cellular response in a specific experimental setup and a different cellular response in another setup (see https://www.nextprot.org for examples). Also and more understandably, the same messenger seems to be able to induce different effects on different cell types.

Contribution of this Thesis

An interesting question is to know if the experimental setup is responsible for the different behaviors observed for the same messengers or if there is a deeper meaning to these discrepancies. For example, there are more and more evidences showing that the nature of the induced response to a molecule is dependent on the internal and external contexts of the cell receiving it. Also we already know that nature optimize resources usage and could have done the same with cell communication systems. Instead of having one messenger molecule for every action a cell can perform, evolution might have selected a system where combinations of messenger molecules carry the relevant information for the cell and not the molecules themselves. Each messenger becomes a word in a sentence but only the full sentence makes sense. This would allow the reuse of the same molecules in different contexts. Is that the reason behind the strange experimental results that are published sometimes?

In order to test this structured communication hypothesis, computer simulation is well suited since it allows a full control of the system environment and should be much easier to setup than *in vitro* experiments. Full cell colony simulation is still a long way away but using suitable hypotheses to simplify the model, very interesting results can be derived from *in silico* experiments. However, when performing these experiments, we need to use methods and algorithms that do not introduce any bias towards the emergent phenomenon we want to investigate. Any method that uses a fitness function to assess the progress of the system towards the expected goal is thus prohibited. This includes neural networks, genetic algorithms and many other popular algorithms. Without introducing any bias, we also need to explore the parameter space of the system in the most efficient way possible since we do not have 4 billion years to experiment. Adaptive multi agent systems (AMAS) are an interesting framework that meets these two expectations. At its core, the AMAS approach uses cooperation between agents to productively sample the parameter space, and its local decision mechanism avoids any possibility of bias towards predefined system wide goals. It is also interesting to note that cooperation in the realm of biology is an essential trait which has made possible the transition from single cell organisms to multicellular "higher" organisms.

On the side of the AMAS approach, this work will explore the notion of cooperation in an environment where agents initially do not have communication protocols to coordinate their actions. It is interesting to see if the principles of cooperation and agent similarity are enough to make a viable communication "language" emerge from the system. Self-organization and adaptation of the AMAS structure are being tested outside of their usual limits and can extend its application scope with the findings of this work.

This thesis describes the development of a simple yet representative enough model of a collection of cells. Using cooperation and kinship, this model is used to explore the potential for resources homeostasis, critical resource sharing and finally emergence of a coordination communication protocol. In parallel to this and since computing power and available time are

limited, cooperation in the AMAS approach is used in order to accelerate the parameter space exploration without introducing any bias.

Manuscript Organization

The manuscript is organized as follows:

Chapter 2: This chapter introduces the biological and computational contexts of this work and the key notions needed to build the cell simulation model. Actual communication mechanisms observed in real multicellular organisms are described, then the state of the art in both biology and informatics is presented. For the biological part, published experiments supporting the hypothesis of this thesis are discussed. For the computational part, software and algorithms dealing with cellular and multicellular simulations are presented.

Chapter 3: In this chapter, CoCell, the contribution of this work, is presented. The choices made to build the simulation are presented before establishing the foundations of its implementation according to the AMAS theory. In this way, components of the simulation are agentified and the roles and behaviors of these agents are introduced.

Chapter 4: CoCell1, the first implementation of the contribution is presented in this chapter. Its performance to maintain a multicellular system alive, as well as its stability and robustness are evaluated through different experiments.

Chapter 5: This chapter enriches CoCell1 with cell mutations to implement the second part of the contribution, CoCell2. The impact of these mutations on the stability and adaptability of the system are evaluated.

Chapter 6: Finally, the last instance, CoCell3, is described in this chapter. This last stage enables the study of the necessary and sufficient conditions to observe the emergence of communication between cells from simple agent behaviors.

Chapter 7: This last chapter concludes and presents perspectives. Implications of the findings of this work on real life multicellular organisms are discussed. Then possible orientations of future works in the area of multicellular simulation are proposed.

CHAPTER 2. CONTEXT AND STATE OF THE ART

Dans ce chapitre, nous discuterons des notions clés en biologie et en informatique afin de comprendre l'hypothèse de communication cellulaire structurée présentée dans ce manuscrit et les façons de l'aborder dans les simulations.

Du côté biologique, le rôle de l'acide désoxyribonucléique (ADN) dans la vie cellulaire ainsi que les comportements de base et les mécanismes de traitement de l'information sont introduits. A partir de cela, l'hypothèse de communication structurée est présentée avec ses implications, les difficultés à la valider expérimentalement et la nécessité de l'aborder à l'aide de simulations in silico. Ensuite, les notions clés pour comprendre la communication entre les cellules sont discutées ainsi que les raisons de l'existence de cette communication et son rôle crucial dans l'émergence d'organismes multicellulaires. Enfin, les caractéristiques essentielles d'un modèle de simulation sont proposées.

Du côté informatique, les obstacles et les limitations potentielles pour simuler les cellules vivantes sont étudiés. Divers algorithmes liés aux simulations cellulaires sont présentés et discutés. Leur pertinence pour la simulation de communication structurée et leur coût de calcul sont évalués.

Enfin, les caractéristiques requises par un modèle cellulaire utilisable dans le cadre de ce travail sont établies.

In this chapter, we will discuss key notions in both biology and informatics in order to understand the cellular structured communication hypothesis presented in this manuscript and ways to address it in simulations.

For the biological side, the role of deoxyribonucleic acid (DNA) in cell life as well as basic behaviors and information processing mechanisms are introduced. From these, the structured communication hypothesis is presented with its implications, the difficulties to validate it experimentally and the need to address it using *in silico* simulations. Then key notions to understand communication between cells are discussed as well as the reasons for its existence like its crucial role in the emergence of multicellular organisms. Finally, essential features of a simulation model are proposed.

For the computational side, potential hurdles and limitations to simulate living cells are envisioned. Various algorithms related to cellular simulations are presented and discussed. Their relevance for the structured communication simulation is evaluated and their computational cost discussed.

Finally, the required features of a framework usable for this work are enumerated.

2.1 DNA is Information but the Rest of the Cell too

Since the dawn of life on Earth, heredity has been essential to maintain kinship and keep favorable traits. Nowadays, we know that the memory that is transmitted from generation to

generation is encoded in a long molecule called DNA[1]. After this finding in the mid-20th century, most of biology focused on decoding the information stored in this molecule and understanding it. Although the full length of the human genome has been sequenced, most of this molecular book of knowledge remains a mystery. Indeed, there are around 3 billion base pairs in the human genome. Around 21,000 genes coding for proteins with biological activity have been identified but this only amounts to 1.5% of the total. The function of the rest of the genome is still subject to debate. Some of the identified roles include regulation of gene expression, epigenetics, genetic interactions, noncoding functional RNA[2], introns, repeat sequences, transposons and viral elements. The interplay of all these elements is still very speculative but one thing is clear: The book of heredity is nothing without its reader. That is, without the complex machinery of proteins, RNA and small molecules that surround and attend to DNA, its information is uninterpretable and useless. Not only that but the nature of the machinery querying the DNA actually changes the meaning of what is "written". This is probably why there are divergent opinions about the usefulness of the Human Genome Project (Evans et al. 2011; Gisler, Sornette, and Woodard 2010). At its start in 1990, there were very high hopes that the full DNA sequence would give us the ultimate insight in the inner workings of the cell and that we would be able to transform the way medicine worked: Drugs on demand, personalized medicine, gene therapy, complete eradication of hereditary diseases and changes of traits were among the promises of this project. Fifteen years after its completion (around 2003), it is not always clear how the sequencing directly contributed to the advances in medicine in terms of new drugs and therapies (Figure 2-1). Maybe this is because we have the book but not the right vision to understand it the way cells do.

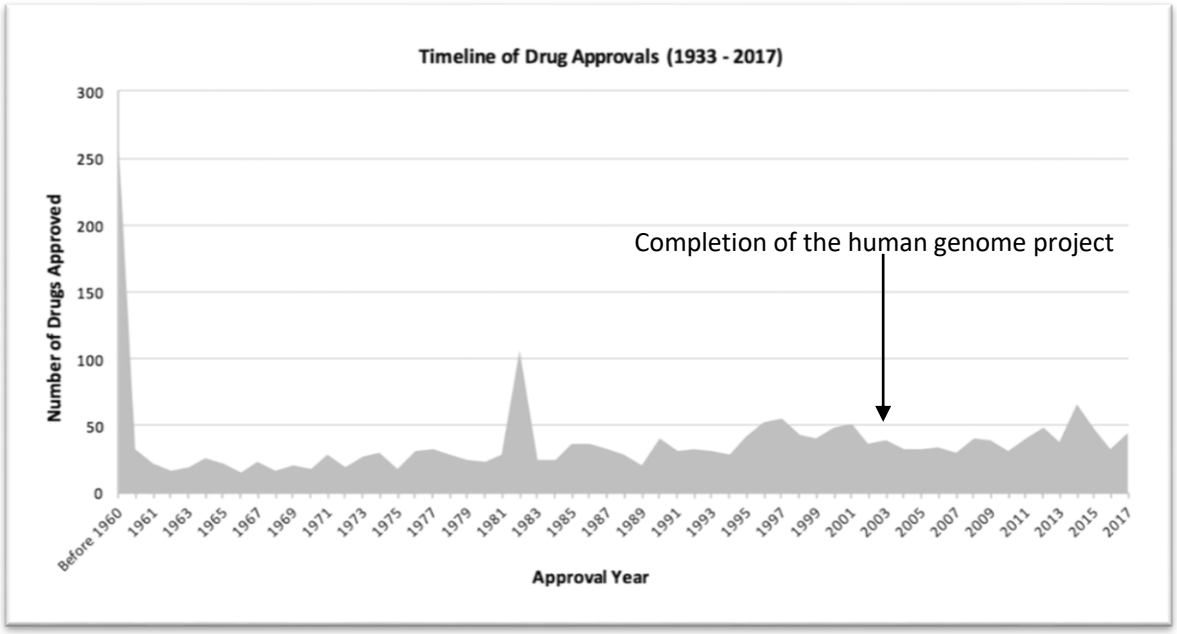


Figure 2-1 Number of new drugs with time (http://cheminfo.charite.de/superdrug2/statistics.html)

[1] This molecule is located in the nucleus of our cells and is dependent of a complex molecular machinery for its translation into actual cell actions.

[2] Ribonucleic acid is usually responsible to transfer genetic information from the nucleus of the cell to "factories" in its cytoplasm (the material within cell, excluding the nucleus) where it is translated into proteins.

As said earlier, DNA without a cell around it is not complete. Hence in order to understand what is encoded, one probably needs to understand what goes around in a cell. At this point, we face a huge challenge: The cell cytoplasm is a complex system. Not complex in the sense complicated because there are a lot of different chemical species, vesicles, organelles and so on, but in the sense that the behavior of the whole cannot be predicted from the properties of the components. The behavior of the cell is an emergent property of its constituents. Knowing the exact composition of the cytoplasm would probably not help us predict its behavior.

To further highlight the importance of the cytoplasm let us consider its function as another important way of storing information in the cell. During cell division, the mother cell splits in two, duplicating its chromosomes to give one copy to each daughter cell. This leads to beautiful illustrations in text books as in Figure 2-2.

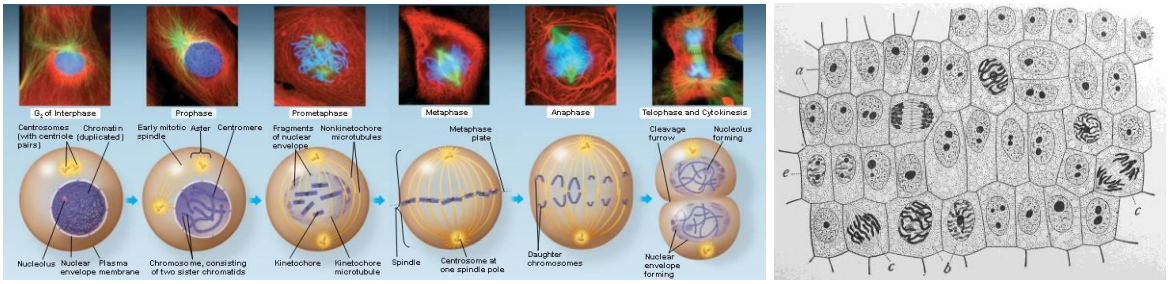


Figure 2-2 Cell division with highlight on chromosome duplication and separation (Pearson Education, Inc. 2011)

But these diagrams and real cell photos are quite biased; they focus the attention on chromosomes (DNA). They do not reflect the underlying duplication of all organelles and other cell machinery. And actually each daughter cell has the same cytoplasmic content as its mother. This is crucial for the new cells in order to start in the best way and be in phase with their environment. Cell specialization for example is "encoded" in this non-DNA memory. The DNA nature and content are the same for a muscle cell, an adipocyte or a white cell; it is the composition of proteins and other molecules that will make a cell what it is. The specific mix of proteins will inhibit expression of some genes and promote others. In turn, these genes will lead to the synthesis of proteins that will reinforce the regulation or modify it. So, without the full system going with it, the DNA knowledge repository does no give much insight on the future behavior of a cell. Furthermore, this cytoplasmic memory can truly be compared to an information storage since it is passed from generation to generation in the same fashion as genetic material. But unlike the genes, it is more dynamic and represents a kind of short term memory.

From these considerations, the future of gene therapy appears much less straightforward than it once was. This could also explain why attempts to introduce new genetic material in a cell has had so little success so far (although inserting the new gene into a cell is also an extremely difficult endeavor).

2.2 Is DNA the Programming Language of the Cell?

Since the discovery of its role in heredity and storage of information, DNA has been considered as one of the best ways to address most of the problems of the cell.

An analogy that is often used is comparing the cell with a computer. Since both are information processing entities it is possible to pair up biological components with electronic components. For example, the membrane is the case of the computer, the mitochondria are the power supply and the cytoplasm is the processor where every computation is performed. Since DNA contains all the

information and the code of what a cell can do and when, it is often compared with the computer programming language. But is it really? As we have explained, the behavior of the cell is not dictated by DNA alone. It is the complex interplay between DNA and cell history (in the form of cytoplasm composition) that influences its future behavior. So introducing new or modified gene into a cell would not really reprogram it if this latter is not in an internal state compatible to accept these new genes. Furthermore, modifying a gene actually changes the structure of the cell. In the computer analogy, it would be the same as modifying components of the motherboard or the hard drive.

If genes are not the programming language of the cell, then is there any and what is it?

A cell, in a multicellular organism, usually does not act on its own accord. To start a new process, produce a new metabolite, differentiate or divide, it needs cues from its environment or direct orders. What are these cues? They can be of several types: Physicochemical properties of the environment like pH, ions concentration, temperature, specific molecules recognized by cellular receptors, gradients of chemicals (protein or small molecules), or physical pressure. Direct orders can come from neighbor cells through signaling molecules or from far away organs through hormones (Figure 2-3). It is actually "quite easy" to make cells divide in a Petri dish: Meet all their resource requirements and then add a growth factor protein at a certain concentration, and they will soon multiply, obeying the request. When an entity responds to stimuli by changing its behavior without affecting its structure, can this be called programming? If the answer is yes, then molecules used in cell-cell communication would be the true programming language of cells. In this context, DNA would represent the blueprint of the machinery that can interpret and execute this programming language. But at this point the cell-computer metaphor is probably reaching its limits since a computer is not a complex system and each of its components has a clearly defined role. In a cell, function boundaries are sometimes blurred and roles are mixed or contextual. For example, DNA stores information but it can also have a catalytic activity and directly transform other molecules (like RNA). An enzyme has a catalytic role but when chemically modified becomes an element of information that can regulate other processes.

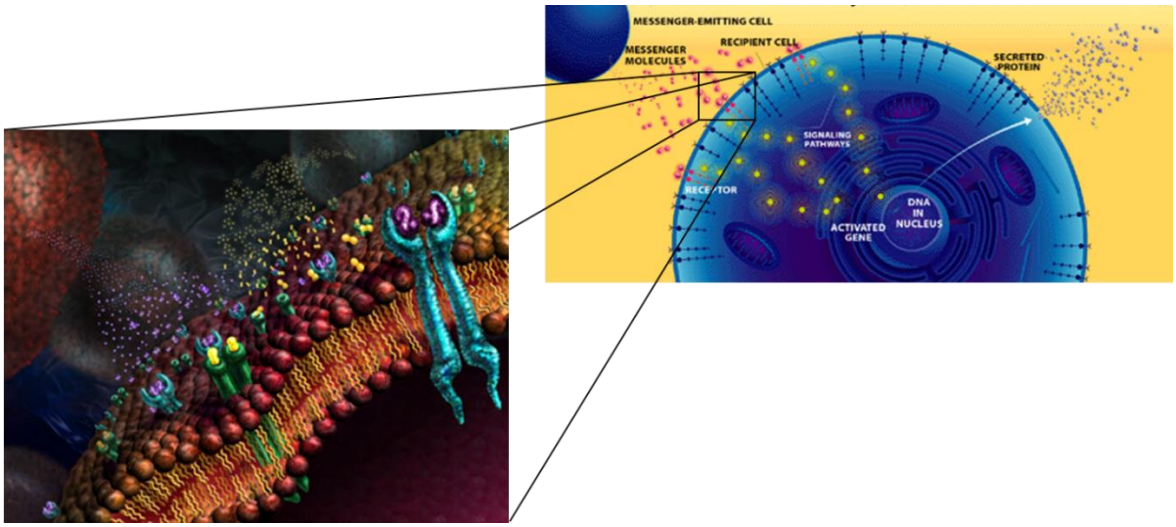


Figure 2-3 Cell-cell communication. The programming language of the cell?

Assuming cell-cell communication can be considered as the way to program cells, several questions arise. a) How exactly does it work? b) Is it possible to interfere with cell-cell communication in order to change the behavior of target cells? c) Could there be practical applications like drugs? d) How to investigate it in greater details?

a) How exactly does it work? Information exchange between cells is more and more understood as the mediators are uncovered (usually proteins) and their effects on various

cell types studied. Basically a signal is emitted by a cell, travels a short distance and binds to a specific receptor on the recipient cell. Upon binding, a cascade of events takes place inside the cell (this is called signal transduction) ending with the change of behavior of the target cell. It should be possible to alter the response of the target cell by modifying the concentration of the signal or by blocking the cell surface receptors for this signal. Obviously in order to do that a big challenge to address is the ability to precisely target the right cells without affecting the others.

b) Is it possible to interfere with cell-cell communication in order to change the behavior of target cells? About half of the current pharmacopeia are molecules that bind cell surface receptors (Figure 2-4). And this trend did not change with the advent of the genome project. These molecules modify in one way or another the communication system of the target cells. So any kind of interference with cell-cell communication has a proven potential to help fight diseases.

At this point another analogy can be used to explain how cell-cell communication can be as powerful as altering the complex internal mechanisms of the cell. In psychiatry, the brain is the complex system to cure from dysfunction behaviors. One way to regulate these inadequate behaviors is to use drugs that will alter the way neurons work. By rebalancing the flux of neuromediators these drugs can alleviate symptoms and sometimes reorient the brain in a more stable equilibrium. In some cases, another way of dealing with disorders is "just" to talk. This method uses the communication protocols of the brain without interfering with its inner workings that are inherently complex (with consciousness as its main emergent property). Communication will deeply alter the state of the brain (its position in its parameter space) and can have the same effects as a drug treatment. Sometimes the effect of such therapies have longer lasting effects than drugs. Usually it is also a much longer process. So the analogy lies between the complex cell represented by the brain, drugs altering the inner structure of the cell, and communication as a way to profoundly alter the behavior of the cell.

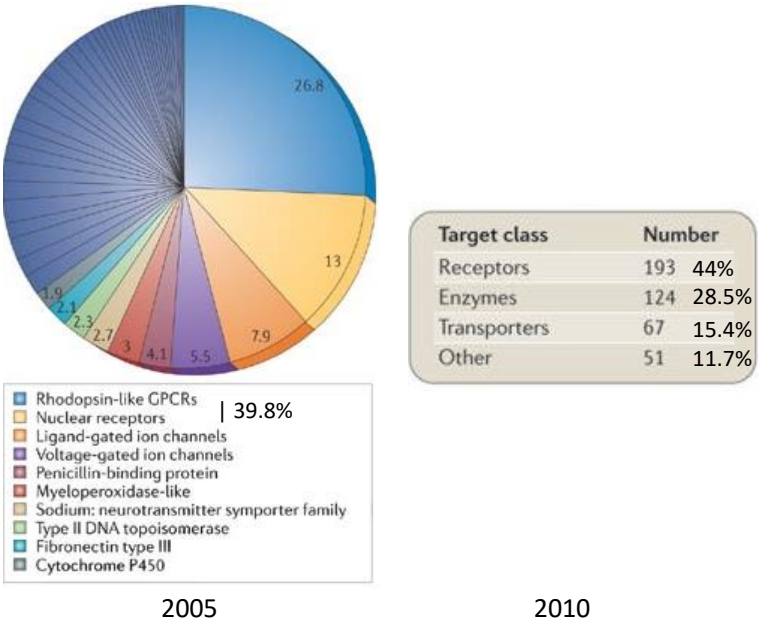


Figure 2-4 Evolution of drug target classes between 2005 and 2010 (Overington, Al-Lazikani, and Hopkins 2006; Rask-Andersen, Almén, and Schiöth 2011)

c) Could there be practical applications like drugs? Using the communication protocols of the cell as a means to change its behavior would potentially have the same results or better

results than the current drugs used to bind cell surface receptors. A huge difficulty would be to target these messages precisely to the specific deregulated cells: The signaling molecules are usually very short-lived in any living tissue and have a potential to activate unwanted behaviors in healthy cells between the injection point and the target site. But this is also true for most small molecule drugs. Furthermore, proteins (that form the vast majority of the signals) cannot be used as oral drugs since they are immediately destroyed in the digestive system.

d) How to investigate it in greater details? In cells it is still difficult to identify the precise role of the different signals. This is due to the fact that the observed effects can vary with experimental conditions and with the cell type under study. New technologies being developed will improve efficiency and precision of these kinds of signals studies. But as in any kind of experimentation the conceptual framework can strongly bias the design of experiments and influence the interpretation of results. In the case of signal biology, the reductionist approach tends to favor studies of single signals at a time and to identify single cellular responses. This can have a strong impact on the deciphering of how this communication programming language works as we will see later on.

2.3 Optimization in Nature: Combination as a Way to Improve Biological Processes

Nature has a proven record for its ability to optimize processes in order to reduce resource and energy consumption. For example, 20 amino acids used in proteins synthesis are encoded in DNA using base triplets called codons. Bases in DNA sequences can be of 4 types (Figure 2-5): A (adenine), C (cytosine), G (guanine), T (thymine). Base triplets give an encoding power of $4^3 = 64$ possibilities (Figure 2-6). So the cell only has to keep the machinery associated with 4 bases types instead of 20 in a one to one coding system. This is clearly more efficient and contains more potential for flexibility, repair systems and future development.

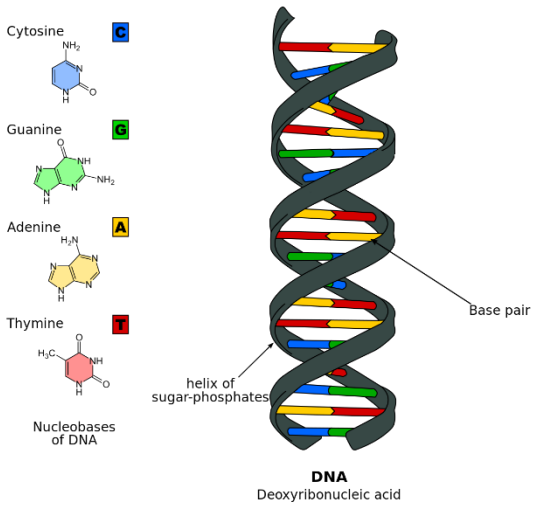


Figure 2-5 The 4 nucleotides used in DNA code

First Position	Second Position: U				Second Position: C				Second Position: A				Second Position: G				Third Position
	code	Amio Acid			code	Amio Acid			code	Amio Acid			code	Amio Acid			
U	UUU	phe			UCU	ser			UAU	tyr			UGU	cys			U
	UUC				UCC				UAC				UGC				C
	UUA	leu			UCA				UAA	STOP			UGA	STOP			A
	UUG				UCG				UAG	STOP			UGG	trp			G
C	CUU	leu			CCU	pro			CAU	his			CGU	arg			U
	CUC				CCC				CAC				CGC				C
	CUA				CCA				CAA	gln			CGA				A
	CUG				CCG				CAG				CGG				G
A	AUU	ile			ACU	thr			AAU	asn			AGU	ser			U
	AUC				ACC				AAC				AGC				C
	AUA				ACA				AAA	lys			AGA	arg			A
	AUG	met			ACG				AAG				AGG				G
G	GUU	val			GCU	ala			GAU	asp			GGU	gly			U
	GUC				GCC				GAC				GGC				C
	GUA				GCA				GAA	glu			GGA				A
	GUG				GCG				GAG				GGG				G

Figure 2-6 Codon table for protein amino acids in humans

Another example of optimization is the use of amino acids in proteins. Instead of custom-made specialized molecules used to perform every necessary task in the cell, the 20 amino acids are the building blocks from which "any" kind of function can be made. Actually this might be the direct consequence of the 4 bases used in DNA since RNA and DNA were present before proteins. Still, it shows that when a process can be made more efficient, time and evolution are able to improve it.

Based on these examples, it would not be too farfetched to imagine that a similar process happened for the communication system. That would mean that a single signal molecule would not be associated with a single response; instead it would be the association of several different signals that would be specific to a single outcome (Figure 2-7). This contextual interpretation of signals, or structured communication, is the core hypothesis driving the work of this thesis. Although it is sometimes admitted that signals interpretation by cells depends on the context, this is not the main trend in current biology and context is often interpreted as the situation of the cell and resources levels rather than the presence of other signals that act as modulators.

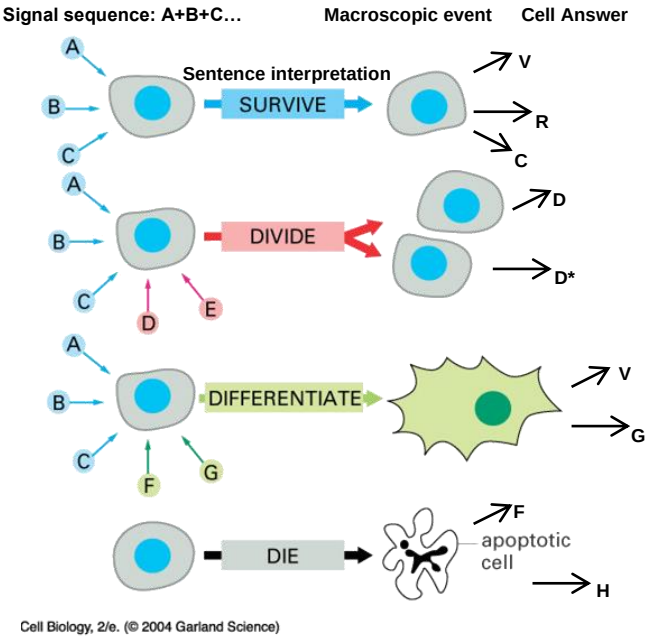


Figure 2-7 Structured communication hypothesis

Several published experimental data support this hypothesis but they are far from numerous (see paragraph 2.8). It may be because these experiments are extremely difficult to design and control or because it is not the current frame of thoughts in biology. Reductionism is the prevailing way of thinking in this field and experiments are designed with this view of the world in mind. This often leads to studies of single signals and the search for single effects. The name of the signaling molecules often reflects this fundamental bias: Tumor Necrosis Factor, Epidermal Growth Factor or Colony-stimulating Factor.

In order to assert or disprove this signal combination hypothesis there are several options. The simplest but most time and resource consuming is to test on cells in vitro, combinations of signals and note the resulting effects (which can be difficult for unforeseen effects). For a set of 10 signals in combination of 3, there are already 1,000 experiments to perform (plus repeats for statistical reasons). The choice of these 10 signals would be in itself a challenge since there are over 200 known signaling molecules. Also the human body counts over 200 different specialized cell types. The selection of the right cell type for the experiment is also important. And even more important is to repeat the experiments on different cell types to note the differences of response to the same stimuli. An added difficulty comes from the fact that cells commonly used in the lab are often immortalized cell lines. This means that they are somehow genetically modified to survive and divide outside the body. These cell lines are much easier to work with since long and multiple experiments can be performed using the same initial batch of cells and they are convenient to obtain in large quantities. On the contrary unmodified cells need to be isolated from donors, are not always numerous, and depending on the cell type can be difficult to maintain alive and/or breed. Sometimes large variations of results can be observed from one batch of cells to another (different donors) which is not suitable for large sets of experiments. But the use of immortalized cells can lead to biased studies since these modified cells do not always behave as their natural ancestors.

A simpler alternative to the pure combinatorial approach consists in testing various signals to try to modulate the already documented main effect of a single signal. If new phenomena are observed this could be used as a bootstrap for further studies since modulator signals (agonists[3] or antagonists) could be tested with other signals to confirm their role. But again, this is experimentally challenging and also this combinatorial approach is not an active field of research.

2.4 Simulating the Emergence of Structured Communication

A third approach, which is used in this work, is to investigate structured communication between cells using cell models and computer simulation. In such a setup, every aspect of the model can be observed. If from a "simple" set of cooperating cells emerges a communication system, one can test if there is advantages to combinations versus a system without it. If the simulation is generic enough to represent the most basic features of living cells, the findings could then be extrapolated to real systems.

The goal of this work is to make a simulation of cells evolve towards multicellularity and to observe the emergence of communication between these cells. If the emergent communication process is based on signal combinations, the hypothesis might be viable. If not, either the hypothesis is wrong or the cell model is no pertinent.

[3] In biology, an agonist is a molecule that activates a cellular response upon binding on a specific receptor. An antagonist blocks this cellular response when binding on the receptor.

The model needs to be carefully elaborated to avoid any bias towards one answer or the other. Furthermore, it must be simple to be generic enough but also representative of the main functions encountered in living cells. Each aspect of the cell functions has to be pondered to decide if it is pertinent to include in the model or not. Also, since the simulation focuses on the emergence of communication, the model has to be closer to primordial cells than to modern cells that already include elaborated messaging protocols.

These reasons require to first have a look at the way a cell works and how it communicates with its siblings. This is necessary in order to make the right choices when building a cellular model. Then available computer simulation algorithms need to be evaluated for their potential to efficiently run a multicellular model with the required features.

There are several aspects of this work that have been extensively discussed in the literature. They fall in two broad categories, biological experiments and theories, for one part, and computer simulations and methodologies, for the other. In the following paragraphs we first approach the biological world at the cellular level and highlight some characteristics that enabled the emergence of multicellular organisms, and in particular communication. From this, we derive the main features that a simulation system must follow in order to represent accurately enough a biological system. Then some of the current cellular simulation platforms and methodologies are presented as well as their relevance to our subject. Computer simulation strategies for complex systems are then discussed and in particular the Adaptive Multi-Agent Systems (AMAS) approach.

2.5 Complex Systems

Since this work mainly focuses on biological cells, it is necessary to describe some of their features. A cell is basically a porous bag of chemicals. This is a very simplistic view of this entity and does not do justice to its potential since it is ultimately responsible for the lines I am writing and for Voyager 1 cruising outside of our solar system. So what transforms a bag of chemicals into such a powerful entity? Essentially it is a phenomenon called complexity and its corollary emergence.

Complexity can be found in many areas of the world around us like in weather patterns, economy, or societies but expresses itself most strongly in the biological field. Definitions vary (Guespin-Michel 2017) but certain features are always present:

- A complex system is usually constituted of several to many components. The components are not necessarily different from one another.
- Components of the system are interacting together.
- Complex systems show emergent behavior. Out of the interactions between the individual elements in the systems, behavior emerges at the level of the system as a whole. This so-called higher order behavior cannot simply be derived from behaviors at the component level. In other words, "The whole is more than the sum of its parts" (Aristotle, metaphysical). This higher order behavior was not "intended" by the elements. It is a spontaneous behavior.
- Complex systems show non-linear dynamics. That means that their behavior may suddenly change. They may move from stable dynamic equilibrium to very unstable behavior. In our daily life we are surrounded by these sudden changes like for example, revolutions and financial crises.
- Relatively small changes may lead to large effects. This is the case if a complex system is close to a tipping point and it is therefore related to the non-linearity of complex systems. Again, these are the result of the inter-connectivity of complex systems components.

Furthermore, if a system is said adaptive, its behavior will change with the environment conditions. The new behavior will provide a better functional fitness with the new external conditions.

A consequence of complexity in a system is that its behavior cannot be predicted with accuracy. Small changes in initial conditions can lead to very different dynamics over time. The existence of non-linear behaviors also adds to unpredictability.

Complex systems exist at different levels of organization that range from the subatomic realm to individual organisms to whole populations and beyond. So it is possible for a complex system to be composed of components being themselves complex systems like it is the case for multicellular organisms or societies.

2.6 Emergence

Like complexity, emergence is a non-trivial term actively discussed. While this notion is central to the study of complex systems, it has no formal definition that is unanimous. Usually emergence happens when a global behavior arises from interactions between the parts of the system under study (De Wolf and Holvoet 2005). What seems important is that the emergent behavior does not exist for any system's components.

There is nearly a definition of the emergence term per field of investigation. From the computer science perspective, emergence is defined as "the process that causes a software system to produce an emergent phenomenon" and its corollary "An emergent phenomenon produced by a software is an interpretation of an attractor the system has converged into, which is practically unpredictable given the functionality of system components" (Di Marzo Serugendo, Gleizes, and Karageorgos 2011).

From the biological standpoint, emergence ranges from the observation of patterns in simple chemical reactions (Dobrescu and Purcarea 2011) to the apparition of self-replicating phenomena commonly called "life".

Examples of emergence are quite easy to find although some of them were not recognized as such for some time. Put two hydrogen atoms and one of oxygen together and you will get a molecule named dihydrogen monoxide, more commonly known as water (Figure 2-8). Even if it is possible to calculate a highly precise quantum description of this molecule and its properties, when you have billions of them together new properties emerge. For example, water boiling point is not readily derivable from the quantum description of a single molecule. This is due to the anisotropic interactions the multitude molecules will form together (Brini et al. 2017). These interactions called hydrogen bonds are essential for the role that water plays in life. Few other known solvents display these interesting properties and could replace it in this role. This might be considered an example of weak emergence (see below).

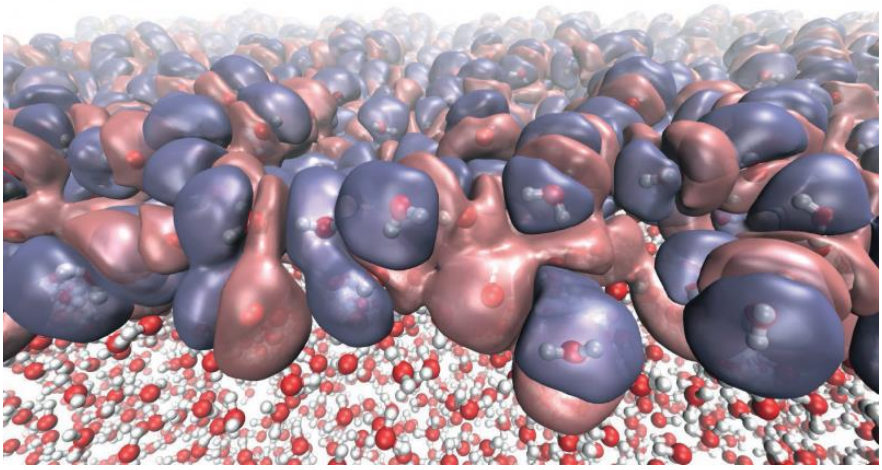


Figure 2-8 Heterogeneous electronic density created by the diverse molecular orientations at the liquid-vapor interface of water (Credit: NPL/University of Edinburgh)

Another well-known example of emergence is human consciousness (Baars and Edelman 2012). Free from any religious interference, consciousness is a direct result of the billions neurons interacting together in our brain. An interesting aspect of consciousness as an emerging property of our brain is that although it can identify oneself as "I" it is unable to determine its own origin or its inner workings. This allegorically shows the transition from the interaction of many simple components to a somehow unrelated collective behavior.

As for many other phenomena, consciousness can also be considered a multi-level emergent behavior. Indeed, for the neurons to interact by electrical signal exchanges, the neuron cell type had to emerge from undifferentiated cells in the body. These undifferentiated cells had to emerge from unicellular cells that started to live as a group. And these unicellular cells had to emerge from "random" chemical reactions in the vast cooking pot that was primordial Earth.

Transition from a micro-level to a macro-level is not enough to define an emergent behavior. Indeed, any system behavior ultimately results from the interaction of its parts, even if the system is not complex like a car. Other criteria need to be taken into account (Goldstein 1999):

- The dynamical aspect: The emergent behavior is not intrinsic to the system. It will arise from the evolution of the system.
- The evolution of the behavior itself: Although the system appears to go back to a state where emergence did not exist, the behavior continues to arise. It acquires an identity by itself.
- The novelty of the features displayed by the system: The emergent behavior has properties not previously observed in the system. These properties are not predictable from the micro-level.
- Decentralization of control: The macro-level is intangible and then not directly controllable by an external entity (such as a supervisor), the control is only possible by the entities at the micro-level. But no entity at the micro-level has a global control on the macro-level.

A fundamental (and highly discussed) aspect of emergence is that it is an ostensible phenomenon. This means that it is only recognizable by showing itself. Recognizable implies that there is an observer to acknowledge it.

An expert with enough knowledge on the behavior of a complex system could unroll step by step the chain of causality between the micro and macro levels. He would therefore not see a

phenomenon as emergent since novelty would be missing. Emergence is a relation between observation and observer. So emergence could be a name for our own limitation to understand and track the multitude of interactions that occur in complex systems at the micro-level. This is called weak emergence.

By contrast, if it is possible to prove that the observed emergent behavior is not the product of a lack of knowledge of the components, it is called strong emergence.

There has been a growing interest about emergence as a design paradigm in artificial systems (Ulieru and Doursat 2011). A key component of this approach is that, although a system can be simple to design, it can exhibit complex functionalities that emerge from the interactions between its parts. So in theory it could be possible to design systems of "simple" components that would exhibit the same behaviors as systems with complicated components using the power of emergent phenomenon. This would allow to identify the essential properties of the component for a given global behavior. This is actually what we try to do by modeling simple cells to observe the emergence of communication.

2.7 Biological Background

Since our focus is on trying to unravel the mechanisms of the emergence of communication between cells using computer simulation, it is important to look how communication works in current biological organisms. From a biological perspective, several questions are of particular interest:

- What is cell-cell communication?
- Is cell-cell communication ubiquitous on Earth or multicellular organism specific?
- What are the prerequisite for communication?
- How did it evolve?
- In terms of natural selection, what are the advantages of communication?

Not all of these questions have clear unambiguous answers. There are still some debate and different theories to account for some observations but general ideas can be derived from these works.

2.7.1 Generalities

Unless noted otherwise, all references in this paragraph can be found in two books, namely (Cooper and Hausman 2009; H. F. Lodish 2013).

Cells are divided into two categories: Prokaryotes and eukaryotes (Figure 2-9). The former cells were the first to appear on Earth around 3.8 billion years ago and are still present today in the form of bacteria. The latter cells appeared 1.5 billion years later (Figure 2-10), possess a nucleus and rely on oxidative metabolism for their energy supply. Eukaryotes are bigger and more complex than prokaryotes; their cytoplasm includes various compartments for specialized metabolic functions. In particular, energy production takes place in mitochondria that are specialized organelles. The fact that these mitochondria possess their own DNA indicates that they once were free bacteria and at one point in the past, became symbionts. Plant cells include chloroplasts which are organelles specialized in the photosynthetic process; again these chloroplasts are symbionts.

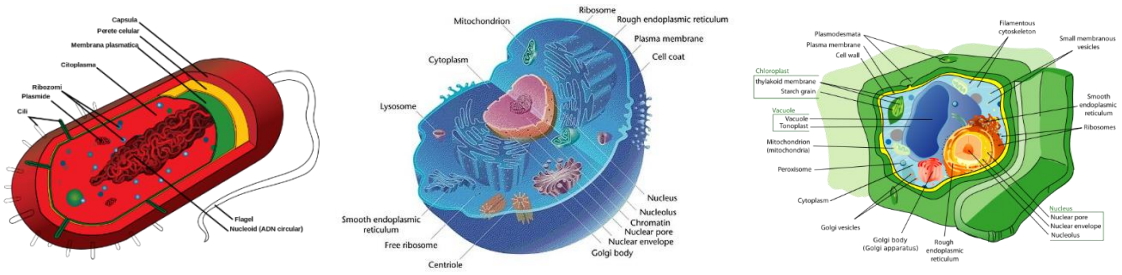


Figure 2-9 Different types of cells: Prokaryotes and eukaryotes (animal cells and plant cells)

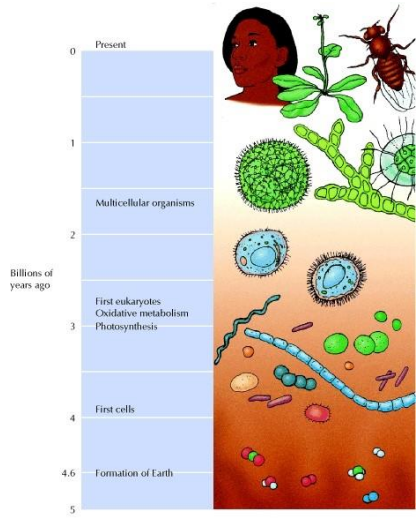


Figure 2-10 History of life on Earth (Cooper and Hausman 2009)

Energy production in prokaryotes is based on glycolysis whereas in eukaryotes it uses oxidative metabolism which requires oxygen but is much more efficient (Figure 2-11). Both mechanisms produce the same molecule to store energy, namely adenosine triphosphate (ATP). This molecule is unstable and is hydrolyzed in adenosine diphosphate (ADP) in water, releasing a large amount of free energy. ATP hydrolysis is used by cells to perform chemical reactions that require high activation energy. These reactions are catalyzed by enzymes (usually proteins) that get together ATP and reactants to enable unlikely chemical reactions. It is interesting to note that ATP and the mechanisms to produce it are very highly conserved across all living organisms on Earth. Although there are numerous alternatives to store energy, only ATP is used by living beings on Earth (Kamerlin et al. 2013).

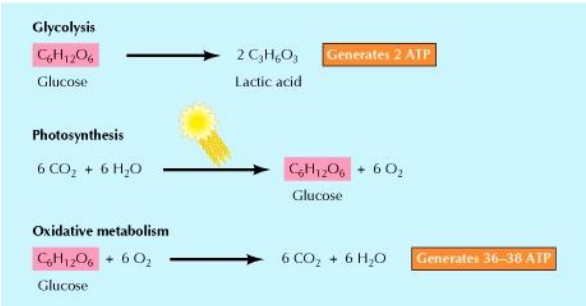


Figure 2-11 Energy production in cells

A cell basically processes chemical resources, transforming them into other chemical resources. This chemical transformation can be spontaneous or highly thermodynamically unfavorable. Evolution selected three means to make difficult reactions possible:

- Separate exterior from interior using a membrane: This is an essential step in the apparition of life as we know it. It allowed the concentration of chemicals to be higher inside than outside which in turn favored chemical transformation by the use of mass action law. In short, this physical law states that a chemical reaction is favored if reactant concentrations is high.
- Use catalysis: A catalyst is a substance that is not consumed during a reaction but will speed it up. In the biological realm, catalysts are often proteins, called enzymes. The role of these macromolecules is to specifically bind the reactants (called substrates) of a chemical reaction and, by doing so, bring them close together favoring the reaction to form the product (Figure 2-12).

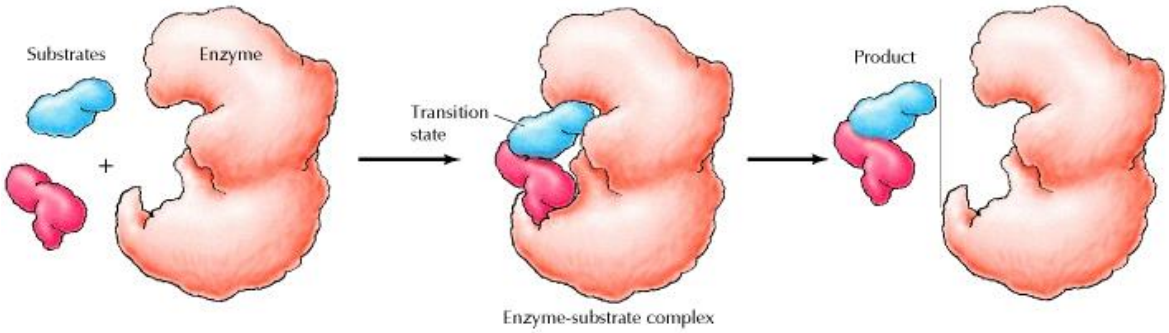


Figure 2-12 Enzyme catalysis

- Store and use energy to speed up chemical reactions: As mentioned earlier, cells use ATP to store energy. This energy can be released on demand in the active site of an enzyme to further help the transformation of substrates to product.

Furthermore, the use of enzyme catalysis provides a powerful added functionality to the system: Regulation.

Regulation is the ability to make a chemical reaction slow or fast based on a criterion completely unrelated to the reaction itself. Enzymes can be activated or deactivated by other proteins or molecules. Feedback mechanisms are a type of regulation essential to the cell. Suppose there is a chain of catalyzed reactions (usually called a pathway in biology) starting from substrate A and leading to the formation of product Z (Figure 2-13). If Z is able to inhibit the catalytic action of the first enzyme in the process, then there is regulation. Indeed, if Z is absent the chain will produce Z unhindered. But as the concentration of Z increases, it will increasingly inhibit the first step in its production. This is a non-linear process that will converge to a stable concentration of Z.

When the enzymes of the Z pathway are inhibited by products or intermediates of other pathways, it gives the cell the ability to have production of Z finely tuned to its need in every possible situation. In a way, this enzyme regulation mechanism is equivalent to transistors in electronics. So, in theory, it would be possible to perform the same calculations in a cell as with an electronic computer. Actually biological systems might even be more powerful since they deal with various chemical entities instead of just electrons. The downside when compared with electronic computers is the maximum reaction speed which is order of magnitude slower than slow transistors. A whole field of biology is interested by this kind of biological computer (Moe-Behrens

2013). The key message from this regulation mechanism is that a cell is able to perform computations and take decision based on available information.

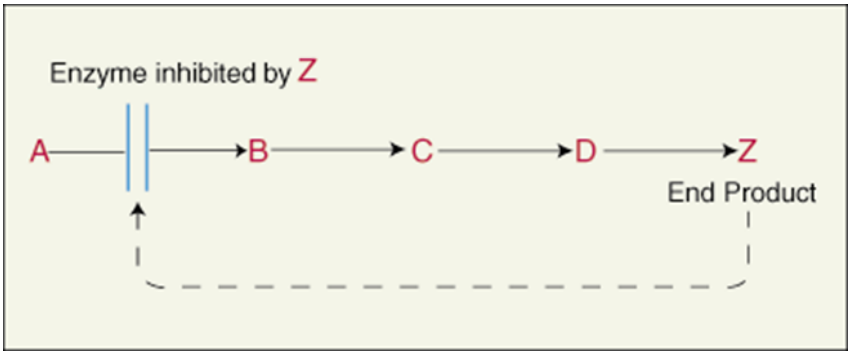


Figure 2-13 Feedback loop example

From the preceding paragraphs, it is clear that a cell not only produces/transforms chemical resources but at the same time processes information. Unlike in a computer, cell constituents that are molecules are both raw materials and information. it is interesting to note that for the most part modern cells have somehow separated information and materials since, for example, signaling proteins like cytokines and growth factors are only used to signal and do not have catalytic activities.

Thus, a cell is able to gather information from its surrounding, process the information, take decisions and then perform an action.

2.7.2 Complexity and Emergence in Biological Systems

In biological systems, complexity gives the impression of following the arrow of time. From the very first cells to nowadays organisms, it seems that everything is more refined, optimized and complex. But this is only apparent and mainly due to a perception bias: We only see the 0.01% of living organisms on Earth that are highly complex like animals, and ignore 99.99% of the small "simpler" life forms that are the microbes.

Emergence in biology is omnipresent at every level either weakly or strongly (Clayton and Davies 2006). Non-linear processes are already present at the lowest levels in chemical reactions. Feedback loops are one of the strongest drive of non-linearity and their apparition is probably one of the first steps to adaptive cells and living beings (Figure 2-14, Figure 2-15).

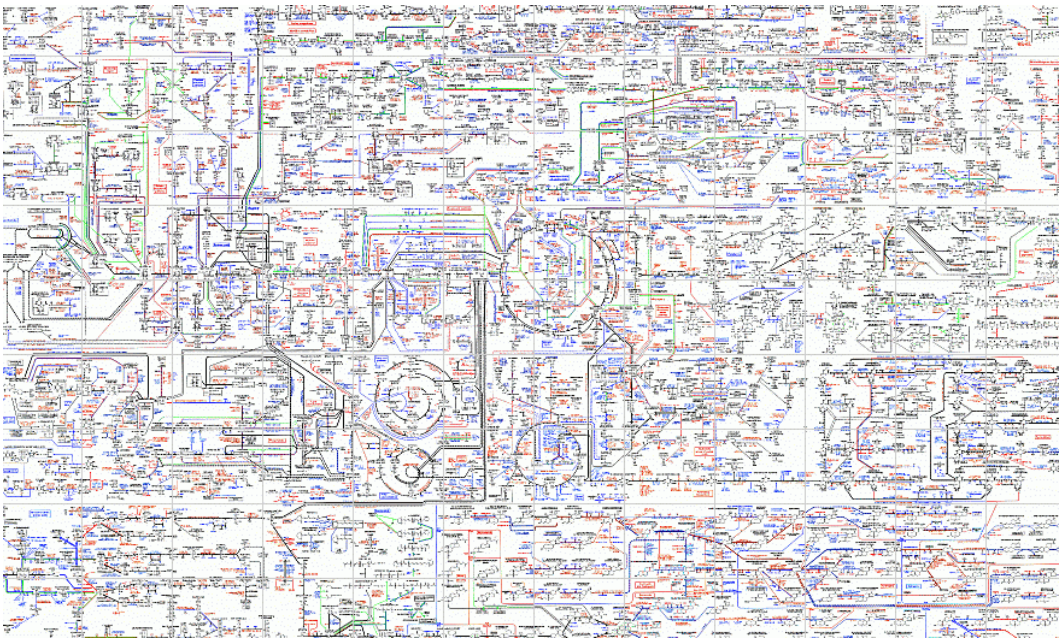


Figure 2-14 Partial view of chemical pathways in a cell. Feedback loops and crosstalks[4] between pathways are present everywhere

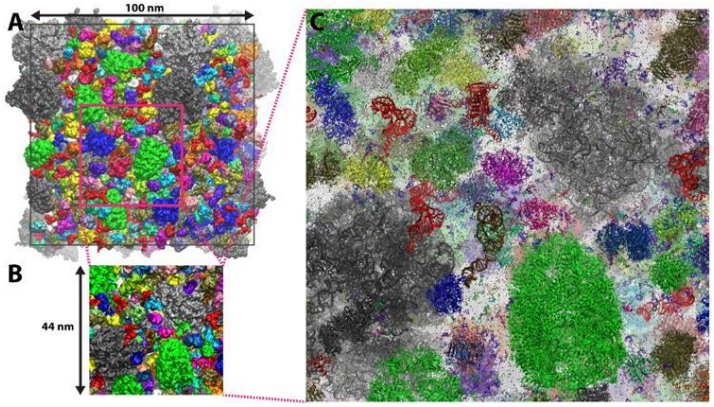


Figure 2-15 Cytoplasmic sample from whole cell simulation of bacteria demonstrating its complexity (Feig et al. 2015)

In fact, biological systems follow a single rule: Only the fittest will live to see another day (Darwin 1859). If complexity can help with that goal, organisms become more complex. If not, they become simpler. The environment will dictate to a biological system if added complexity is a bonus or not. When an organism is competing with a virus that hijacks its complex reproduction mechanism, either it adds more defense systems to block the virus or it "reverts" to a more basic life form that does not rely on the hijacked processes (or it disappears…). If resources are plentiful, the defense systems that would add to resource consumption are a viable solution. But if it is already difficult to find enough resources to sustain the organism as it is, added complexity would probably prove fatal and simplification is the best option.

Biological systems use an adaptive mechanism often referred to as "evolution" to survive changing external conditions. New features will emerge from this process and render an organism able to survive in a different environment. The main drive of this process are mutations that create

[4] Crosstalk in biology indicates that pathways are able to influence each other. Products of one pathway can inhibit or activate another pathway and reciprocally.

novelty and the second drive is selection that will filter out useless novelty. This selection process is quite simple since for the mutating organism it is either survival or death.

Adaptability in Nature is the result of the "domestication" of an irremovable phenomenon: Random noise and its biochemical corollary, mutations. Random noise exists at nearly every scale in the physical world. From quantum fluctuations at the atomic level to explosion of faraway supernovae that can swipe clean the surface of a planet light years away. This noise can be intrinsic to the nature of phenomena, the result of combinatorial explosion or sensitivity to initial conditions in a deterministic system. The extraordinary feat of biological organisms was to turn this wild unstoppable force of chaos and entropy into a strength. Repair systems in cells are very efficient and so are DNA duplication mechanisms; but only up to a certain point. If they were too perfect (which is actually impossible) no mutations could happen and the system would only be able to survive in an unchanging environment. By controlling the amount of mutations that are allowed to appear during a period of time, an organism can introduce novelty in its behavior in order to adapt to a changing environment and at the same time preserve its identity.

Mutations in a real cell can happen anytime during its lifetime. Their appearance may have different causes: Environmental influence through chemicals or ionizing radiations (UV, X-rays, gamma rays), failure of the DNA replicating mechanism (estimated to be 1 per 10 billion base pair) or spontaneous chemical modification. There are several types of mutations:

a) Silent mutations where the mutated codon still codes for the same amino acid, hence the resulting protein is the same. There may be long term influence of silent mutations since the various codons for the same amino acid do not have the same preference in a given organism (this is called codon usage). That is, a codon A for an amino acid can be translated faster or with more accuracy than a codon B for the same amino acid. Thus, the dynamics of the system can be subtly altered with unforeseen consequences. In non-coding DNA, and as far as we know, silent mutations do not change the function of the segment.

b) Neutral mutations are similar to silent mutations since they do not have a direct effect on the mutated protein because the change of amino acid does not impact the shape or function of the protein.

c) Harmful mutations change a crucial amino acid codon in a protein (like a catalytic amino acid) or an important base in a regulatory non-coding DNA sequence. The effects are immediately damaging to the function of the cell and usually end up with the cell death. More deleterious mutations can have more subtle effects than let the cell survive but endanger the organism as a whole like oncogenic mutations than affect cell proliferation and can translate into a cancer.

d) Beneficial mutations can affect either a regulatory non-coding DNA or a protein gene. In the former case, regulation of some genes is altered, improving the fitness of the cell in its current environment. In the latter case, the mutated protein may be more efficient in its functions or may be able to perform something entirely new and required to improve the fitness of the organism.

In a multicellular organism most mutations affect cells that will not pass their genetic material to the next generation. This are called somatic mutations. Some cells are more sensitive to mutations than others and will "prefer" apoptosis (cell suicide) than performing an altered function that could endanger the whole organism or the species. For example, germline cells that are involved in the transmission of genetic information to the offspring have a mutation rate ten times lower than somatic cells (Milholland et al. 2017). This shows that mutations transmitted to the next generation are strictly filtered, because potentially dangerous for the species. This also means that

they can be controlled and contained to a certain extent. Thus, cells have turned an unavoidable entropic decay into a strength to adapt.

Evolution is not a directed process. There is no centralized control designing specific changes to enhance organism survival. Chance mutation is a blind mechanism that takes a lot of time and trials to find a viable answer to a problem. This random walk in the space of possible changes is not very efficient but it is not a problem when there are billions of years to search and trillions of organisms as test subjects.

2.7.3 Emergence of Multicellularity

Multicellularity appeared "recently" in Earth history, about 1.7 billion ago. It is important to note that this phenomenon is not unidirectional *i.e.* once multicellularity appears it is always possible to go back to single cell organisms. Actually it is thought that multicellularity appeared and disappeared several times in the past (Duran-Nebreda et al. 2016; Grosberg and Strathmann 2007; Niklas 2014; Ruiz-Trillo et al. 2007). This would suggest that increased complexity and evolution are not linked. As mentioned before, sometimes and because of changes in the environmental conditions, a decrease in complexity can be beneficial and may give a competitive edge to simpler organisms.

Some advantages of multicellularity include:

- Decreasing the risk to become a prey by increasing the size of the organism (D. L. Kirk 2003).
- Resistance to physical and chemical stresses (Justice et al. 2008): The list of stresses against which multicellularity affords a defense includes temperature, pH, osmotic pressure, oxidation, desiccation, metal toxicity, and mechanical forces.
- Generation of an internal environment protected by an external layer of cells (Lyons and Kolter 2015).
- Allowing novel metabolic opportunities (Zhang, Claessen, and Rozen 2016): Division of labor allows multicellular organisms to be more efficient in terms of resource production and management.
- Enhanced motility for dispersal or foraging, expanding feeding opportunities (H. Koschwanez, R. Foster, and W. Murray 2011): Cells in groups can better take up extracellularly-produced resources that would otherwise diffuse away, a tactic that single cells would find both inefficient and susceptible to freeloaders.
- Providing storage reserves when nutrients are limiting growth.
- More efficient colonization of new territories.

Some disadvantages of multicellularity that could lead back to unicellular organisms are:

- Bigger size means small organisms can invade: Although size represents a way to escape predation it becomes a disadvantage when smaller organisms are able to invade, leading to infectious and parasitic diseases.
- More energy is needed for normal functioning: Multicellular cells are usually more complex than unicellular ones, and this is often correlated to more energy consuming processes and higher resources requirements.
- Takes longer to reach maturity and to breed: To duplicate the delicate layout of cells is always more time consuming than the basic cell division used by unicellular organisms. For example, it takes 9 months for humans to "half-duplicate" whereas it takes 20 minutes for the Escherichia coli bacteria.

- If one cell group fails, they can all fail: In multicellular organisms, cells are usually very specialized and need other cell types for their survival. If one cell type disappears for one reason or another it can endanger the whole organism.
- Susceptibility to cheaters: Some cells can revert to a living mode where they profit from the common resources but do not participate. In this mode, cheaters do not spend as much energy as other cells and can thrive. Sometimes, this can lead to the death of the multicellular organism. An example of cheaters are cancer cells, although this particular kind of cell does not minimize its energy consumption but maximizes its duplication rate (Aktipis et al. 2015).

It is not completely clear how multicellularity emerged and if it is weak or strong emergence. Nevertheless, several key requirements have been identified in order for unicellular organisms to forgo their freedom (Grosberg and Strathmann 2007; David L. Kirk 2005):

- Cell adhesion (Driscoll and Travisano 2017): Cells developed the capacity to stick together either by incomplete division or by specific mechanisms. This is a trivial necessity in order to become a multicellular entity. Nevertheless, this is not as straightforward as it would seem since once glued together a cell must also resist the external pressure of its peers. The simultaneous apparition of an internal cytoskeleton is thus necessary (Jacobeen et al. 2017).
- Specialization, division of labor (Flores and Herrero 2010): Some metabolic processes are chemically incompatible and cannot occur at the same time in a single cell. Cell specialization is a possible way to perform incompatible tasks at the same time at the multicellular organism level (Figure 2-16).

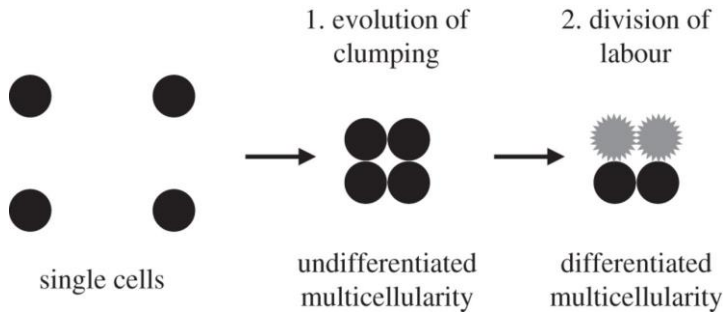


Figure 2-16 Two key steps in the major transition to multicellularity (Biernaskie and West 2015)

- Programmed cell death (Claessen et al. 2014): The death of some cells can serve the interests of the greater community, for example to provide extra nutrients or create raised structures that increase surface area and thus nutrient exposure.
- Ratcheting system to favor multicellular vs unicellular (Libby et al. 2016; Libby and Ratcliff 2014; Lukeš et al. 2011): All necessary mutations leading to an efficient multicellular organism are unlikely to appear simultaneously. In order to accumulate the necessary mutations without reverting to unicellularity, a ratchet system is necessary to keep evolving towards multicellularity; for example, interdependencies between specialized cells or decreased probability that a mutation will result in reversion.
- Cell-cell communication and coordination (Diggle et al. 2007): Once cells are specialized, they need a way to request what they do not produce anymore and to know when neighbor cells lack critical materials. A chemical communication system is then necessary to keep the multicellular level working.

The last item is of particular interest in this work. More specifically, how can this communication emerge from dependent cells and how is it structured?

2.7.4 Cell-cell Communication Systems

Progress in molecular biology, cell biology and observation technology during the last 30 years allowed to dissect the inner working of cells and, in particular, the way they exchange information.

2.7.4.1 Prokaryote Communication

Until recently, prokaryotes where thought to be lonely organisms fighting for survival in a sea of competitors. Actually this is not the case, and communication, even if much simpler than for eukaryotes, does exist in the realm of bacteria; it is called quorum sensing (Miller and Bassler 2001). Basically, each bacterium in a colony produces a molecule specific to its species (called autoinducers) and monitors its level in the medium. When this level reaches a certain threshold, several physiological responses are triggered. Since the autoinducer level is proportional to the number of active bacteria in the colony, it will enable colony level decision-making. For example, virulence factors that are directly linked to host invasion are only released when the colony is strong enough; otherwise, the host immune system would be triggered too soon and would kill the few aggressive bacteria.

2.7.4.2 Multicellular Eukaryote Communication

Unlike prokaryotes, multicellular eukaryotic organisms are fundamentally dependent on their cell-cell communication system. One of the main reasons is that division of labor creates interdependencies between cells and their survival is linked to their capacity to request key resources to their neighbors. Thus it is not surprising that communication protocols are multifaceted and much more elaborate than for prokaryotes.

There are multiple ways eukaryote cells communicate together (H. Lodish et al. 2000) :

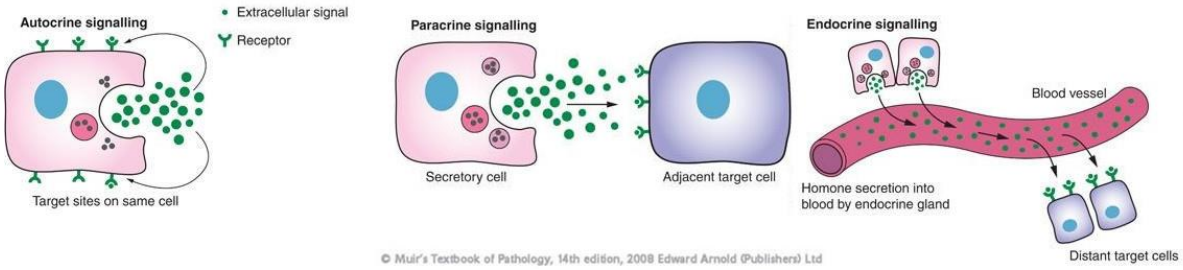


Figure 2-17 Extracellular communication modes

- Autocrine communication (Figure 2-17).
  - The recipient of the message is the emitter itself. This is often observed in immunity response cells.
- Paracrine communication.
  - The emitter cell secretes messenger molecule(s) in its immediate vicinity. The signal is perceived by adjacent cells through specific receptors.
- Endocrine communication.
  - Long distance communication using the bloodstream to transport the messages (hormones) far away from the signaling cell location.
- Juxtacrine communication.

- A special case where the signal molecule stays bound to the emitter cell and binds to a receptor on the receiver cell. In this case, cells must be in very close contact to communicate.

- Direct contact between cells: Gap junctions (Figure 2-18).
  - Protein channels allow free passage of small molecules between cells: These small molecules are direct information sharing.
  - These junctions can be regulated to allow passage of large molecules.

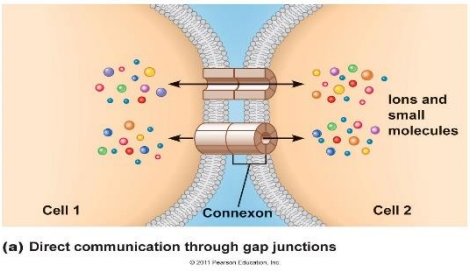


Figure 2-18 Gap junctions

- Chemical modification of the Extra Cellular Matrix (ECM)
  - The signaling cell will modify the ECM in some way.
  - Other cells will be able to detect these modifications and change their behavior accordingly.
- Synaptic signaling is performed by neurons and is both electrical and chemical (Figure 2-19).
  - The electrical signal is sent along the neuron axon.
  - At the end of the axon is located the synapse where the electrical signal triggers secretion of a chemical that acts as a neurotransmitter.
  - Although there is information transfer between the neurons, this information is not used by the cells to change their internal behavior but is only useful at the emergent level of the computational machine that is the whole brain.

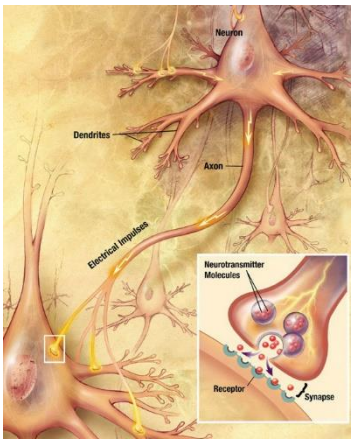


Figure 2-19 Neuronal communication

If the emitter cell displays receptors associated with the signals it releases, autocrine is a side effect of paracrine. In this sense, by simulating a simple system we will *de facto* observe autocrine communication.

Endocrine is a very interesting system for long distance communication, but there lacks evidence that a cell can send multiple hormones to a destination organ and that they will arrive at the same

time on the target cells. Hence the endocrine system may not be relevant for the signal combination hypothesis of this study. Furthermore, simulating a long distance system without first simulating the local communication of cells would seem like to put the cart before the horse.

For the rest of this work, we will only focus on the paracrine communication system, as it appears to be the dedicated system to locally coordinate communities of cells and is ubiquitously used in multicellular organisms.

2.7.5 Paracrine Signalling

Over the last decade, genetics helped identify and classify genes coding for receptors and signaling proteins. About 200 signaling molecules (mostly proteins) (Ben-Shlomo et al. 2003) and their associated receptors have been found so far (http://www.receptome.org). They are categorized into several families that correspond to their genetic relationship, chemical nature or structure. Classification by the type of induced effect(s) on cells is not possible since these can be different depending on the target cellular type (Gaudet et al. 2017).

Most of these signaling molecules are stored in vesicles inside the cell cytoplasm and released in the environment when the cell needs to send a message. Only cells displaying a specific receptor on their surface are susceptible to receive the information: They are called competent cells. This generally results in the activation of second messengers inside the cell, leading to various physiological effects.

Signaling molecules in the paracrine system have usually short half-life (the time required to reduce the molecule concentration to half its initial value) which limits their reach to 3-4 cell layers (~100µm) around the emitter cell (Handly, Pilko, and Wollman 2015; Thurley et al. 2015). This is a key parameter to have a local communication system with maximal efficiency and reduced signal/noise ratio.

Notable exceptions to this rule are nitric oxide (Bryan, Bian, and Murad 2009), hydrogen sulfide (Kimura 2015) and carbon monoxide (Kim, Ryter, and Choi 2006). Indeed, these volatile gases can freely diffuse across cell membranes and they have very short half-life. They act as messengers in diverse functions including vasodilation, neurotransmission, anti-tumor and anti-pathogenic activities. They are not stored and there are no receptors to detect them. Their mode of action is unusual because of their ability to cross membranes. Once they have been synthesized they diffuse isotropically in the cell cytoplasm and then in the extracellular environment until they enter neighboring cells to induce a response.

2.7.6 Summary of Biological Features Important for Simulation

In this section, we have learned that cells are small production plants where chemical resources are transformed using energy extracted from ATP molecules. More than that, cells have become information processing units able to adapt to a changing environment. Survival (as ATP production) and reproduction are the main focus of cells and more generally life as we know it.

Multicellular organisms heavily rely on communication in order to survive in a very competitive environment. Division of labor and interdependence are strong drives to induce communication and cooperation among cells.

In order to study the emergence phenomenon of interest for this work, a simulation model needs at the very least to take into account:

- The environment as the only means of interactions between the cells;

- Resources and their diffusion in the environment;
- One or several energy resources;
- Difference between cell interior and exterior for the intake and release of resources;
- Resource reactions to form products;
- Cell processing power;
- Cell division;
- Cell death;
- Cell evolution: Ability to vary its capacity to perform some resource reactions.

2.8 Experimental Support of the Structured Communication Hypothesis

The true test for any theory in natural sciences is experimentation. Sometimes experiments are straightforward providing a simple yes/no answer, and sometimes experimental errors and/or unknown experimental parameters render interpretations of the results highly difficult. Sometimes experiment setup takes minutes, like measuring the pH of a solution, and sometimes it can take years like detecting gravitational wave. There is a third case where experiments are not possible given the current limitations of our technologies and knowledge, and in this case it may still be possible to test some hypothesis by using (computer) simulation.

Combinatorial testing and practical experimentation do not go well together. Even very few parameters with few possible values each quickly lead to intractable numbers when combined. This is often referred as combinatorial explosion. An example of such a study is available in (Natarajan et al. 2006) where 22 signals where studied alone or in pairwise combinations on a single cell type. Higher order combinations were not studied because of combinatorial explosion.

Several methodologies have been designed in order to reduce the exponential growth to a linear one. A commonly used approach is fractional factorial design (Jaynes et al. 2013). It allows to efficiently sample the parameter space to drastically reduce the number of required experiments to optimize the response curve. The main problem is that to apply this process a strong assumption about the system response is necessary: It has to be (more or less) linear with each parameter. Synergies, antagonisms or strong dependencies between parameters are also bad for an efficient design. Unfortunately, this approach cannot be used efficiently when looking at cell behaviors. The problem is that we are looking for behaviors resulting from different combinations of signals, and these behaviors might not be related together in any way.

In the special case of biology, an interesting approach can be used to test combinations of molecules (Kainkaryam and Woolf 2009). It consists of pooling the molecules to be tested in mixtures of several compounds. These mixtures are then tested. If a response if observed it can have several causes since there are more than one combination in the pool. Nevertheless, by designing experiments in a way that any given combination of two compounds is only encountered a few times in the tests then the observed effect can be unambiguously attributed to this unique combination: This is called deconvolution. The only strong assumption to use this methodology is that positive response to pools are a rare event. Otherwise deconvolution is not possible since too many combinations could be responsible for the observations.

An alternative when possible is to carefully select the parameters to combine, using prior knowledge to select relevant combinations and dosage. Although this method can reduce the number of experiments to a manageable size it is heavily biased and has the potential to miss new/unpredictable phenomena.

Although these methodologies can help deal with combinatorial explosion, experimental data in this field are quite uncommon when compared to single signal studies. Nevertheless, some notable examples convey interesting results.

2.8.1 Neovascularization Therapy

The first one deals with tissue regeneration and in particular neovascularization (G. Krenning et al. 2013; Guido Krenning, van Luyn, and Harmsen 2009). The process of blood vessel creation is very delicate and requires several steps like cell recruitment, differentiation and proliferation. The number of potential applications is very broad and intense research is focused on this issue. Progenitor cells are cells that can differentiate into a number of specialized cells often required in a complex process like neovascularization. They have shown promise to promote and orchestrate this process. But using these rare cells bring problems of their own. In particular, the harvesting of these cells and their incorporation on the site of neovascularization. Because of this, the authors focused their effort on the identification of paracrine exchanges between the progenitor cells and the tissue cells. They identified a pentad of signals (namely IL-8, MCP-1, HGF, bFGF and VEGFa) that were sufficient to induce the same neovascularization process *in vitro* and *in vivo* than the progenitor cells themselves (Figure 2-20). They avoided the problem of combinatorial explosion by selecting signals based on their observation and extensive knowledge of the process under study.

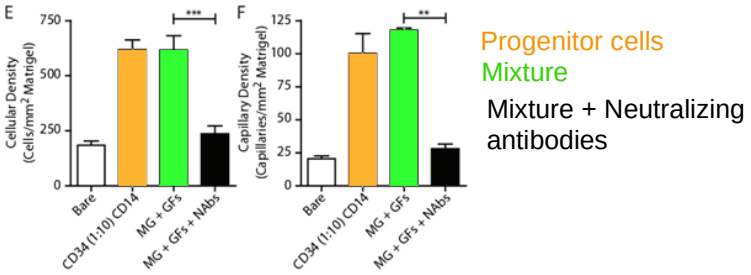


Figure 2-20 Comparative effect of progenitor cells and paracrine signals on neovascularization (G. Krenning et al. 2013)

These results would suggest that a combination of signaling molecules has the desired effect to change the behavior of cells. However, since the process is multistage and involves several cell types it is possible to wonder if the paracrine factors combination acts in different ways for each stage or if signals are acting individually on specific cell types.

In the same field of investigation but with a different setup, (Kwon et al. 2014) found that a similar triplet of paracrine signals (namely IL-8, MCP-1 and VEGF) is necessary and sufficient to induce angiogenic activity *in vitro* and *in vivo*.

2.8.2 Reprogramming Malignant Cells into Normal Cells using Paracrine Signals

In this example (McClellan et al. 2015), BCR–ABL1^{+} B-ALL cells have a malignant phenotype and the authors studied their reversion into normal cells that resemble normal macrophages and can perform macrophage-associated functions. The combination used for the reprogramming included cytokines interleukin 3 (IL-3), M-CSF, granulocyte/monocyte (GM)-CSF, Fms-related tyrosine kinase 3 ligand (FLT3L) and interleukin 7 (IL-7). After treatment of cancer cells from 12 different patients they observed phenotype reversion in 7 cases. The treatments and observations were performed *ex vivo*. Reversal of phenotype was based on loss of invasiveness and leukemogenicity, gene expression profiles and various cell surface markers. As in the case discussed in 2.8.1, the choice of signals to combine was derived from previous observations for each of the individual signals and

extensive data on the pathology under investigation. But unlike neovascularization, phenotype reversion involves a single cell type so it is indeed the signal combination that influences the behavior of cells.

2.8.3 Cytokine Combination Protect from Viral Infection

In (Hartmann et al. 2014), the authors were interested in elucidating the mechanism behind infection resistance in human dendritic cells (DC). Whenever DC cells are infected by the Newcastle disease virus they emit paracrine signaling to warn uninfected cells around them. Upon receiving the message, these cells will modify their behavior to prepare defenses against the infection. The nature of the message was studied and after analyzing signal molecules emitted by infected cells, 20 cytokines were identified as potential candidates. Subsequent experiments identified IFNβ, TNFα, and IL1β as the major contributors to this warning message. Individual cytokines were not able to generate the defensive state in cells. Only the combination of the three paracrine molecules could transmit the right message both *in vitro* and *in vivo* (in mouse).

2.8.4 Calcium Signaling Associated with Pairwise Agonist Scanning

Intracellular calcium concentration is an important parameter of platelet[5] signaling in the clotting process after blood vessel injury. Measuring the influence of extracellular signals on calcium concentration is key to understand the succession of events leading to clot formation. This in turn can enable prevention strategies to deal with heart attacks and strokes.

Six different agonists of human platelet response were tested alone and pairwise combination to identify potential synergies or novel behaviors (Chatterjee et al. 2010). These selected agonists are convulxin (CVX; GPVI activator), ADP, the thromboxane analog U46619, PAR1 agonist peptide (SFLLRN), PAR4 agonist peptide (AYPGKF), and PGE2 (IP receptor activator). 64 triads of ADP, SFLLRN, and CVX were also tested. Pairwise scanning consists in testing individually all possible combinations of two agonists among the six selected on cells and observing the induced cellular response.

A neural network (NN) was then trained to predict calcium concentration using only the agonist pairwise results. The NN successfully predicted the results of the triads of agonists (correlation coefficient $R>0.85$). It was then used to calculate the full 6-dimensional response of mixtures of agonists. Prediction for 45 combinations of 4, 5 or 6 agonists of particular interest were then experimentally tested and found in agreement with the NN values ($R>0.8$).

The conclusion of the study is that pairwise measurements are enough to predict the behavior of higher order combinations. This means that the combination of two signals dominates the response spectrum of platelet cells and that measurable effects of combination of more signals are rare.

The authors agree that their findings cannot be generalized since platelet response presents some specific characteristics as the intracellular wiring that rapidly converges on calcium regulation, without the possibility of higher order effects from genetic regulation or other interactions on long time scales. Furthermore, the study was solely focused on calcium concentration measurement and therefore novel cell behaviors could not be detected.

[5] Platelets are specialized blood cells that help blood clots formation to stop bleeding and start tissue regeneration.

Nevertheless, the notion that binary messages could represent the bulk of cell-cell communication is of great interest although diminishing the potential interest of signal combinations.

2.8.5 Macrophage Cytokine Release

A full combinatorial analysis of cellular response to combinations of signals is quite a feat because of combinatorial explosion. Consequently, such studies are rare in the literature. One of them was published by Hsueh and colleagues (Hsueh et al. 2009). Macrophages (RAW264.7), which are cells of the immune system, were subject to stimulation by combinations of IL-6, TGF-β, IFN-β, isoproterenol and 8-Br-cAMP. After stimulation the secretion of six factors (G-CSF, IL-6, IL-10, MIP-1α, RANTES and TNF-α) was measured. The choice of ligand was driven by the focus of the study on TLR4 receptor activation during sepsis[6]. Similarly, the measured factors were selected for their role in the immune system response during infection. The main finding of this study is that pairwise combinations of signals have different ligand-induced cytokine secretion profiles but higher order combinations do rarely induce unexpected profiles. As in the previous study, the conclusion suggests that "sentences" of chemical signals are often very short. Nonetheless, like in the previous study, the focus of the authors was limited to a specific range of macrophage behaviors (TLR4 modulation). The signals used are all known to modulate TLR4 activity and their measure of profile difference is based on synergy coefficients.

As in 2.8.4, pairwise signals appear to dominate the cell to cell communication with few meaningful messages with more than two "words". But these two studies are very focused on the selected signaling molecules and in the cellular responses. The goal of these studies was more to find synergies between agonists of the same pathways than new cell behaviors. In this light they cannot be considered as definite proof that long message sentences are always meaningless.

2.8.6 Monocytes Impacts the Cytokine Environment

In (Schrier et al. 2016), the authors focused on the impact of coculture of CD4+ T cells and monocytes (white blood cell) on intercellular communication. CD4+ T cells are mature T helper cells that help/regulate the activity of other immune cells like monocytes by releasing T cell cytokines. In this experiment, cocultured cells were subject to three types of stimulation, and 48 different cytokines concentration were simultaneously measured to record cellular stimulations. In the case of two communicating immune cell types, the majority of cytokines were altered from the value measured on isolated population under the same stimulation conditions. This would tend to prove that response from one cellular type to a stimulus will impact the response of the other cell type in a non-additive way. Combination of signals are therefore important for the final cellular action in this model.

2.8.7 Conclusion on Experimental Biology

It appears that literature on the topic of inter cellular communication is often focused on the impact of a single signal on cell behavior. When trying to find broader scope experiments dealing with context dependence of multiple signals, the number of references drops dramatically. From the few papers that can be found, it is difficult to prove or disprove the structured communication hypothesis, but they can contribute to the argumentation. What arise from these experiments is

[6] Sepsis occurs when chemicals released into the bloodstream to fight an infection trigger inflammatory responses throughout the body. This is often a life-threatening condition.

twofold: In experiments dealing with multi-stage processes, communication between cells involves many signals that are necessary and sufficient to complete the function. Still, in this context, it is not clear whether the multiple signals are aimed together at the same cell targets or individually at various cell types. The second finding is that combination of signals known to modulate the same cellular function do not have significant different outcome than single signals when more than two signals are used. The setup of these experiments somehow limit the scope of this finding since the signal used are very specific or non-natural.

2.9 Computational Methods

When real world experiments are too difficult, too long, dangerous or expensive, a valuable alternative is numerical simulation. Obviously simulation requires a reasonable amount of knowledge about the system under study in order to be able to design a suitable model. Apart from the model, efficient algorithms are also necessary to optimize the computational resources. In the case of the structured communication hypothesis we have seen that experiments are faced with many hurdles among which combinatorial explosion is the main one. These hurdles can be minimized in a simulated environment using the proper models. In this chapter, models, simulation frameworks and software dealing with cellular simulation are presented and discussed.

2.9.1 Models

To seek to establish a set of rules that govern the behavior of a complex system raises the issue of the model. In science, the model is a central concept; it is therefore defined in several ways, depending on the domains and the authors. In this work, we consider a model as a simplified representation of a part of the world (Bousquet, Le Page, and Müller 2002). In this perspective, a model can be seen as a set of rules capable of predicting the behavior of the said system (at least in the short term). In this sense, it does not express the reality of what the observed system is, it is just a functional representation. Note that the notion of model is very close to the notion of theory, and the frontier between the two is sometimes difficult to draw (Frigg and Hartmann 2012). Having complex systems models opens up many perspectives like demonstration, decision support, teaching and so on. Even if the scope of applications of such models is vast, we will focus on the aspect that motivated our work: Computer models and simulation of biological cell colonies to understand the emergence of communication behaviors.

2.9.2 Computer Simulation

Why be interested in simulation? Four major areas of science can be identified (Winsberg 2015): Validation of models and theories, prediction of data from a model, understanding of studied systems and representation of data.

Once a good model is available for a system, one way to simulate it is to physically build another system based on the model but in a controlled environment or at a different scale than the original. For example, a tide simulator of 900m^2 has been built to understand how the Mont St-Michel bay would evolve with time. In biology, to study cell mechanisms it is often easier to use model cells in an artificial but controlled environment than real organisms *in vivo*.

The alternative to physical simulation is computer simulation. The advantage is that all parameters can be controlled. The caveat is that models of the simulated system need to be quite accurate and processing power is often a limitation for large or complicated systems. A computer simulation makes it possible to explore the behavior of a system (physical, mathematical, biological,

etc.) by visiting its different successive states from a time t. For this, the simulation uses the state of the system at step x_t to calculate the new state of the system at step x_{t+1}, and so on. The calculation of these new states is done using the model of the simulated system, to calculate the transitions.

As mentioned before, in some fields of science, the current technology and physical limitations do not permit direct experiments. For example, string theory is predicted to have a visible influence on experiments performed at energies several order of magnitudes above the current state-of-the-art particle accelerators. In this case, simulation can provide insights of the predictive power of a theory or of its shortcomings. The analysis of the simulations can yield some predictions that could possibly be tested in real world experimentations.

Simulations can help predict the behavior of a system under various conditions and particularly in the future. A good example is weather forecast for which models are quite good but calculating system states are extremely computationally intensive.

Understanding a system is also easier in a simulation since each state can be thoroughly analyzed and parameter correlations observed. From this, new theories can be developed.

Representation of data derived from a simulation can be used to reinterpret events in a human meaningful way. Teaching is a good application field for this branch of simulation.

Computer simulation is therefore a field of application for which the possibility of having complex system models is important. The challenge is to be able to generate these models, and to validate them.

2.9.3 Models and Techniques Used in Cell Simulation Systems

Simulation in biology is of great interest since many types of experiments are impossible to perform in the laboratory given available technologies. As for any kind of simulation in science, various goals can be achieved: Testing scientific hypothesis, planning experiments or interpreting experimental results. In the particular case of cell simulation, it can also help elucidate the connection between emerging properties of complex systems and micro-scale simple rules.

There are several methodologies to simulate a biological system depending on various parameters. Models range from discrete- to continuous-time, deterministic to stochastic, non-spatial to spatial and can consider single of multi levels of organization. The granularity needed for the description of the biological problem is one of the most important parameter to select the proper type of methodology to use. At one end of the spectrum lies the "exact" simulation of all physical phenomena occurring in the system. At the other end, most of the physical properties are encapsulated into some sort of high order mechanism or equation that emulates more or less correctly the real system at a much smaller computational cost. However, approximations need to be carefully pondered in order to introduce as little bias as possible in the simulation outcome.

Then there are various modelling formalisms to choose from in order to implement the simulation. Some of the common ones used in biology are agent-based modelling (An et al. 2009), boolean networks (Kauffman 1969), Bayesian networks (Wilkinson 2006), cellular automata (Deutsch 2016), constraint-based modelling (Becker et al. 2007), Cellular Potts models (Graner and Glazier 1992), interacting state machines (Kugler, Larjo, and Harel 2010), membrane systems (Barbuti et al. 2011), ordinary/partial differential equations (Hoops et al. 2006), Petri nets (Hardy and Robillard 2004; Heiner, Gilbert, and Donaldson 2008), process algebras (Feng and Hillston 2014), discrete event simulation (Kurve, Kotobi, and Kesidis 2013) and rule-based modelling

(Nikolić, Priami, and Zunino 2012). Several of these methods are reviewed in (Ji et al. 2017) for their use in a biological context.

An important aspect of cell modeling and simulation is multi-scaling (Meier-Schellersheim, Fraser, and Klauschen 2009). Multi-scale simulation generally refers to mathematical and computational models that simultaneously describe processes at multiple time and spatial scales (Figure 2-21). In contrast to the models based on the *quasi-steady state* assumption that discard interactions between scales, multi-scale models describe systems where processes at different scales can influence each other. Therefore, these models should not only simultaneously describe multiple scales, but also allow them to interact. One fact that complicates this is that processes at different scales are often best expressed in different modeling formalisms. Therefore, multi-scale modeling often also involves coupling different modeling techniques that may include spatial/non-spatial models, discrete/continuous models, and stochastic/deterministic models.

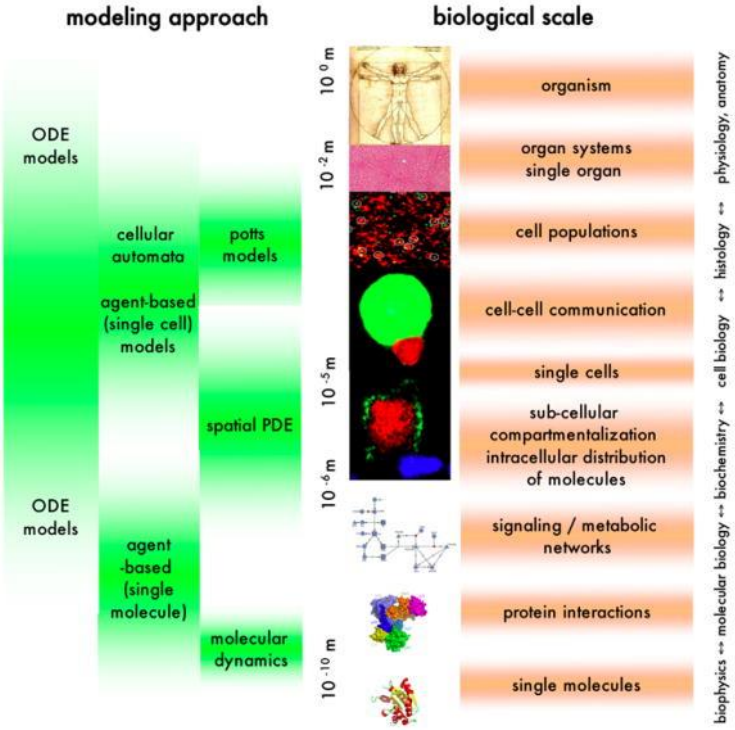


Figure 2-21 Representation of biological scales and their associated modeling techniques. Abbreviations: ODE: Ordinary differential equation; PDE: Partial differential equation (Meier-Schellersheim, Fraser, and Klauschen 2009)

Some of the models used in biological simulations are presented thereafter. Their strengths and weaknesses are discussed. It is important to keep in mind that they are not mutually exclusive and more often than not several ones are used together.

2.9.3.1 Whole Cell Models

The "Holy Grail" of biological simulation would be to simulate a system at the atomic level or even better at the quantum level (Carrera and Covert 2015; Goldberg et al. 2018). Data, knowledge and computing power are very far from allowing to reach such a goal.

Using standard molecular dynamics simulation, it can take from hours to days to simulate one second of a single protein life in water (Adcock and McCammon 2006). A typical mammalian cell contains about 8e9 proteins (Sims and Allbritton 2007), and even more various small molecules like ATP, sugars and so on (without counting water molecules). Even with the most powerful computers available to date it is far from possible to perform an atom-based simulation of a whole cell.

Actually, computation power is maybe the least of the problems. Precise and verified data about the nature and content of a cell is a huge challenge. It is true that experimental techniques are improving by the day and that it is now possible to follow a single protein during its life in the cell. Nevertheless, we are still missing a lot of crucial information and as we know, complex systems can be very sensitive to small changes. Furthermore, at the molecular scale, observation means perturbation and it becomes difficult to obtain unbiased data. Even if in the future we are able to identify all the molecules present in a cell and their physicochemical properties to simulate them, there will still be the huge difficulty to compile a valid starting point for the simulation. Indeed, the dynamical state of a cell is very important and identifying the position of each molecule at a given time on a single cell is a challenge far beyond any foreseen technology.

Until we progress on these key areas some compromises have to be made to even attempt to simulate the simplest whole cell (Figure 2-22).

An example of such an endeavor is bacterium *Mycoplasma genitalium* (Karr et al. 2012). Over 900 sources were used to build the processes taking place inside the cell. No atomic description of the molecules was attempted and ordinary differential equations (see 2.9.3.5) were used to solve the variations of molecule concentrations over time. The simulated model was validated against several types of experimental observations, and was used to provide insight on several processes and provided interesting predictions.

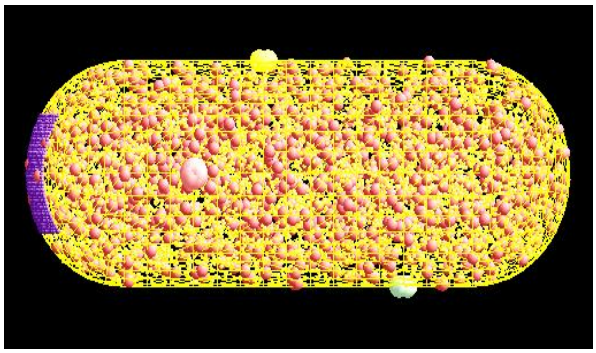


Figure 2-22 Example of "Whole cell" simulation of E. Coli represented in ChemCell (Lipkow, Andrews, and Bray 2005)

2.9.3.2 Atomic-level Models

Atomic-level models are closer to physics than biology. The approach consists in taking into account all atoms included in the system and evaluating, at each time step, the forces applied to them (Figure 2-23). Then using classical Newtonian physics, accelerations are derived from these forces, then speed and positions. In more advanced simulations, when atoms are close together and given their reactivity, new bonds can be created. When they are too far apart, bonds may be broken. This is the standard approach of molecular dynamics (Adcock and McCammon 2006). Several hypotheses can be made in order to speed up the exploration of parameter space and to focus a simulation on interesting events like in the metadynamics approach (Laio and Parrinello 2002). Nevertheless, as already mentioned earlier and given the number of atoms even in small proteins, these models are usually restricted to very small systems.

The advantage of these models is that no assumptions are made about the role of any molecule in the system. So no bias is introduced. Drawbacks of the atomic-level simulation are that it requires three-dimensional knowledge of each molecule present in the system and their respective concentrations. Also, computation power and time grow very fast with the size of the system being simulated. Commonly, this approach is used when studying protein structure dynamics, protein-ligand binding or protein conformational changes.

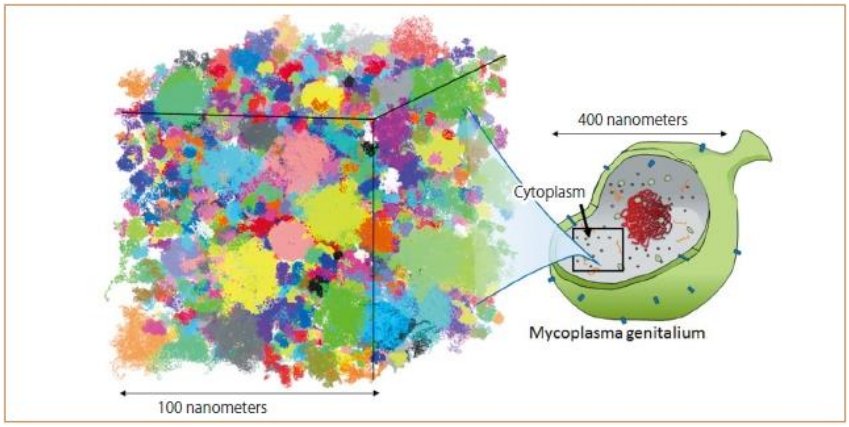


Figure 2-23 Protein content (blobs of various colors) of a 100nm cube (about 100 million atoms) from cytoplasm simulation (Feig et al. 2015)

For multicellular models where billions of atoms are present, the atomic-level approach seems unlikely to be useful even using strong simplifications.

2.9.3.3 Particles-based Models

Particles-based models are in-between full rule-based models and atomic-level description. They use accurate physical laws (like diffusion and collisions) applied to proteins and molecules, but these entities are considered as point-like particles in a continuous space and each has characteristics derived from experimental data (Andrews et al. 2010) (Figure 2-24). This representation is a powerful tool to represent the formation of gradients and cell polarization at a reasonable computational cost. There is also an advantage when compared to differential equations since to be valid these equations require strong assumptions like a well-mixed medium (Alves, Antunes, and Salvador 2006). Using this level of details in a model allows to observe phenomena like morphogenesis where local concentration of molecules is key.

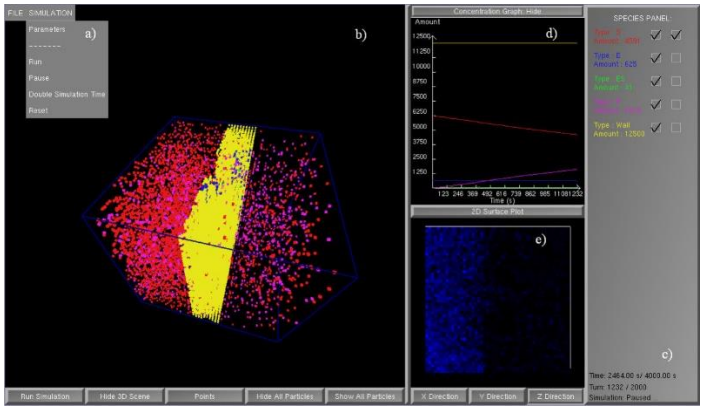


Figure 2-24 Translocation of particles through a membrane with embedded enzymes (Boulianne et al. 2008)

This kind of models can be advantageous to paracrine communication investigation if the number of cells in the system is not too large. Nevertheless, even in a small system this approach is much costlier than a rule-based approach (see section 2.9.3.10) and would become beneficial only if paracrine molecule release is considered anisotropic.

2.9.3.4 Cellular Potts Models (CPM)

In multicellular simulation, a widely used approach is the cellular Potts model (Graner and Glazier 1992; Szabó and Merks 2013; Voss-Böhme 2012). This is a lattice-based, multi-particle cell-

based modeling approach. It is well suited to deal with interactions between cells. The basic model consists in assigning each node of the lattice to a cell or to the medium. Then at each time step simulated, a random grid point at the surface of a cell is selected to attempt a move on a neighboring empty grid point (Figure 2-25). The probability for the success depends on the change in constraints for the cell represented as an energy function called the Hamiltonian. This Hamiltonian can be modified to take into account various influences like cell-cell adhesion strength, chemical gradient in the medium and so on. The granularity of the lattice and the complexity of the Hamiltonian strongly influence the accuracy of the simulation and the computational power required.

Figure 2-25 Cellular Potts model of a tumor. Each color represents a type of mutation present in the tumor

2.9.3.5 Ordinary Differential Equations Models (ODE)

Ordinary Differential Equations (ODE) are based, as their name suggest, on differential equations describing the variation of the concentration of molecules during a small time step. The concentration variation of one species can depend on the concentration of another. Thus, it is possible to have multiple equations describing various species and they are all solved for each time step of the system. These models are of particular interest to describe the evolution of molecule systems with time (Gardner, Cantor, and Collins 2000; Gratie, Iancu, and Petre 2013). They are continuous models and can give predictions for any time value (Figure 2-26). One limitation of this approach is that it cannot take into account random events (however, a modified version – Stochastic Differential Equations – can). Furthermore, ODE do not allow to calculate concentration of a species at different locations in the simulation space. It is interesting to note that ODE do not involve any physical representation of the objects to simulate. So they represent the counterpoint to atomic-level models or particles-based models. At the same time, they are less generic since they require some knowledge on the time dependence parameters and species relationships. Generally, ODE systems are suitable for modeling small-scale networks, since there are many parameters that need to be estimated. If the network scale is large, parameter estimation will lead to high computational cost, and the prediction accuracy of the model may decrease.

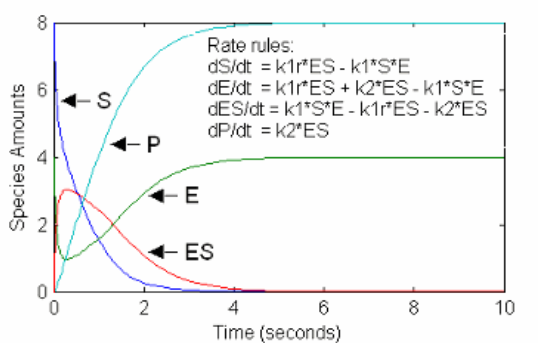


Figure 2-26 Various species evolution with time using ODE

2.9.3.6 *Partial Differential Equation Models (PDE)*

As for ODE, partial differential equations models are continuous models describing variations of molecular concentrations (Clairambault 2013). The added value of these models lies in their ability to calculate species concentration in time but also in space (Figure 2-27). When dealing with morphogenesis where gradient of signaling molecule concentration is key to the phenomenon, PDE becomes a tool of choice.

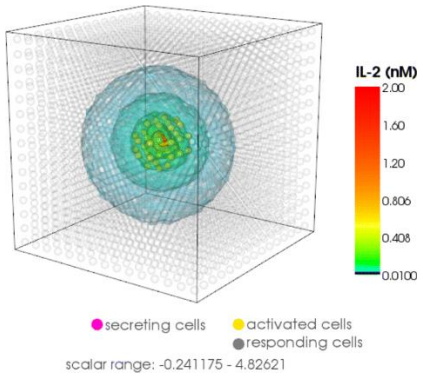


Figure 2-27 IL-2 cytokine diffusion model using PDE (Friedmann 2015)

2.9.3.7 *Lattice Gas Automaton (LGA)*

An alternative to ODE and PDE to simulate the diffusion of molecules is the Lattice Gas Automaton (LGA) (Frisch, Hasslacher, and Pomeau 1986). In this model, particles on a grid move from node to node depending on their speed and direction, and particle collisions are taken into account. Particles change direction and velocity by conserving density and momentum. Although the implementation of such a model is relatively easy and the computational cost is quite low, it comes with several drawbacks: Actual density at a particular grid point is easy to calculate but is very noisy, only average on large parts of the grid give reasonable results, and this model is difficult to adapt to 3D grids.

For these reasons, this model is seldom used and with modern computers, ODE and PDE are the preferred methods.

2.9.3.8 *Petri Nets*

A Petri net (PN) is a directed, weighted bipartite graph consisting of two types of nodes: Places and transitions (Tomar et al. 2013) (Figure 2-28). This type of description is well suited for modeling metabolic pathways, signaling pathways, gene regulation or chemical reactions. During transition

firings, the source places transfer a number of tokens to the target places. The weight of a transition indicates the minimal number of tokens in the sources to enable the transfer. The initial number of tokens in each place defines the starting state of the system. The evolution of the system is determined by the transitions that may occur at each simulation step given the tokens present in the places.

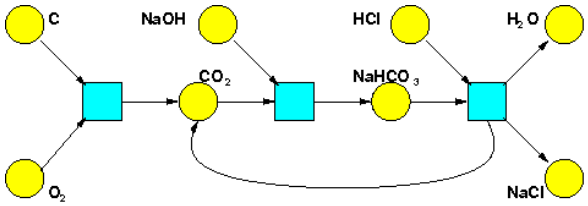


Figure 2-28 Simple Petri net example. Places are yellow circles and transitions are blue squares

As for ODE, using Petri nets requires a precise knowledge of the pathways to simulate, the interactions between them and the feedback loops. In some cases, missing data can profoundly alter the results as usually the simulations deal with chaotic systems very sensitive to initial conditions.

2.9.3.9 Stochastic Models

ODE and PDE are deterministic models than enable the precise calculation of various species concentration present in the system. But in biology, there are often external events that will modify the system or phenomena for which space and time dependencies are not well understood and thus difficult to translate into equations. Another source of randomness lies in the nature of chemical reactions themselves: Equations are valid when dealing with statistically large groups of molecules, but become increasingly inaccurate when the number of elements becomes smaller and probability of encounter between reactants is low. This leads to seemingly random fluctuations of various parameters of the system which may strongly influence its behavior.

One way to take into account these events is to fall back to atomic-level or particles-based models. But as mentioned before these models are computationally costly and quite unsuitable for multicellular simulations (at least for now).

Another approach consists in modifying the ODE model by adding noise to the driving equations of the system. These new equations are called Stochastic Differential Equations (SDE) (Ditlevsen and Samson 2013) .

Monte-Carlo methods are also very popular to simulate certain aspects of biological models where some part of randomness is necessary. These methods are based on event probability. The scientist describes the system in terms of possible events that can take place and their chances of happening. Then during the simulation, a random number generator is used to draw events according to their probability. Since this approach is very generic, it would be quite difficult to summarize all of its uses in biological simulation. (Raychaudhuri 2013) present some of these applications.

2.9.3.10 Rule-based Models

Most often than not whole-cell simulations or even particles-based models are not needed to address biological questions. A partial or coarse-grain model can be enough to observe interesting features of the system. Validation with experimental results is always key to confirm that simplifications in the model do not have too much impact on the phenomenon being investigated.

One such coarse grain simplification is the rule-based model (Hwang et al. 2009). In this approach, chemical reactions are explicitly written as well as their associated key parameters like dissociation constants of reaction speed. Other rules can also be implemented like "if the concentration of protein A is above threshold Ta then gene Ga is activated". Then, this set of rules is executed step after step allowing molecule concentration to change and events to happen in the system. This approach is much simpler than an explicit model of all the atoms present and the use of their reactivity during collisions to perform chemical reactions. Nevertheless, it requires some prior knowledge of the system to establish the relevant rules.

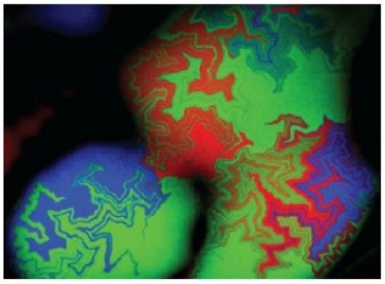


Figure 2-29 Rule-based mechanisms exploiting natural bacterial colony growth can lead to fractals (Scholes and Isalan 2017)

The rule-based models are prevalent in biological simulations. And either by mixing it to other types of models or by itself, this approach is well adapted to communities of cell simulators. In the case of cell-cell communication simulation this type of model can prove very powerful by simplifying several aspects of cell life.

2.9.3.11 Cellular Automata Models

A cellular automata model is a special case of rule-based model since it uses rules but also includes a representation of space and usually deals with a large number of cells (Vivas, Garzón-Alvarado, and Cerrolaza 2015). Space is represented as a grid, where each position is occupied by a cell. Each cell possesses a set of rules that dictates its behavior. This approach is interesting when the biological problem involves more than one cell. It is very popular in the world of complex systems since very simple rules can generate a system with emergent properties. Nevertheless, by itself this model lacks a key component for biological studies: There is no explicit representation of the environment. To fully develop its potential, it has to be merged with other types of algorithms to take into account key aspects of biological systems (Figure 2-30).

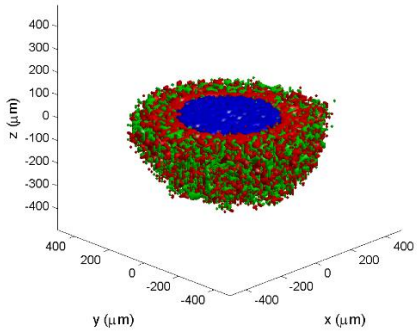


Figure 2-30 Tumor spheroid using cellular automata simulation in BioFVM (Ghaffarizadeh, Friedman, and Macklin 2016)

2.9.3.12 Genetic Algorithms

Unlike their name would suggest, genetic algorithms are optimization processes. The name is inspired from biological natural selection (Forrest 1993). They can be used in biological simulations where more than a single cell is present in the system and when long-term evolution is of interest. There are three important steps in the process: Mutation, crossover and selection using a fitness function. The crossover part is not essential and depends on the simulation scenario. This optimization algorithm is very efficient in very diverse areas of science.

For example, in a cell population simulation, each cell is represented by a set of genes that encode its possible behaviors. The fitness function is defined as the survival potential of a cell. When a cell divides, each daughter cell genome is randomly mutated allowing new behaviors to appear in the population (mutations). Some mutations will represent an advantage for the survival of the cell and others will prove lethal in the conditions of the simulation.

The main problem for cellular simulations is that novelty is random. So if combinations of mutations are required for the behavior of interest to appear in the system, computational time may quickly become prohibitive. As mentioned at the beginning of this chapter, random mutation and survival of the fittest is a powerful tool either if the problem is not combinatorial, or when sample size and allowed time are huge.

2.9.3.13 Multi-Agent Systems (MAS)

Multi-Agent Systems (also called Agent-Based Models (ABM) in biology) are physical or computerized systems defined as an environment shared by a collection of interacting and autonomous entities, the agents. Each agent has only limited information about its environment. The MAS approach is a methodological way to study complex systems with a bottom-up approach. MAS are used in many different scientific areas, from collective problems solving to the study of collective behaviors. To perform a specific task, the MAS methodology proceeds by designing the agent level in order to generate a global behavior that fulfils the request. The distribution of elementary tasks inside MAS makes them particularly suited to address greater complexity than the complexity apprehended by conventional methods. Situated MAS are well-suited for multicellular simulations since each agent could represent a cell. An added feature is that agents can have behaviors different from one another. This approach is very flexible since agent internal workings can be of any kind: Rule-based, neural network or even a complex simulation of cell cytoplasm. For these reasons this is a very popular framework (Börlin et al. 2014; Mina, Tsaneva-Atanasova, and Bernardo 2016; Walpole et al. 2015) (Figure 2-31). Key concepts behind Multi-Agent Systems are presented in section 2.12.

Figure 2-31 Cellular agent-based model (An et al. 2009)

2.9.4 Summary of Required Computational Features

To investigate the emergence of communication between cells and to test the structured communication hypothesis the minimal features required for a computer simulation are the following (accompanied by usable approaches):

- A lattice-based simulation with multiple cells and cell types located on this grid: Cellular automata;
- A diffusion algorithm for resources in the medium: PDE, SDE;
- Division, differentiation, mutation algorithms: Genetic algorithms;
- Autonomous decision-making for cells: Multi-agent systems;
- Chemical reactions inside the cells: ODE;
- A process to accelerate the evolution of the system.

2.10 Software for Cellular Simulation

All the methods described in 2.9.3 have been used alone or in association in numerous software. It is not possible to present here all the software developed around biological simulations. There are nearly as numerous as there are biological questions. Some popular platforms or dealing with multicellular simulation are presented here with their strengths and weaknesses. Their potential use in the context of this thesis is also discussed. They are presented in chronological order.

2.10.1 Netlogo

NetLogo (Wilensky 1999) is a multi-agent programmable modeling environment which was not designed specifically as a biological tool but its structure matches quite well a multicellular setup in an environment (https://ccl.northwestern.edu/netlogo). Over the years, it found many applications in the biological simulation world (Caccavale et al. 2017) (Figure 2-32). To use it, it is necessary to learn its language that is specific to multi-agent programming. It is sufficiently generic to enable the implementation of various algorithms. Nevertheless, it can be tricky to mix several level of simulations or algorithms in this language.

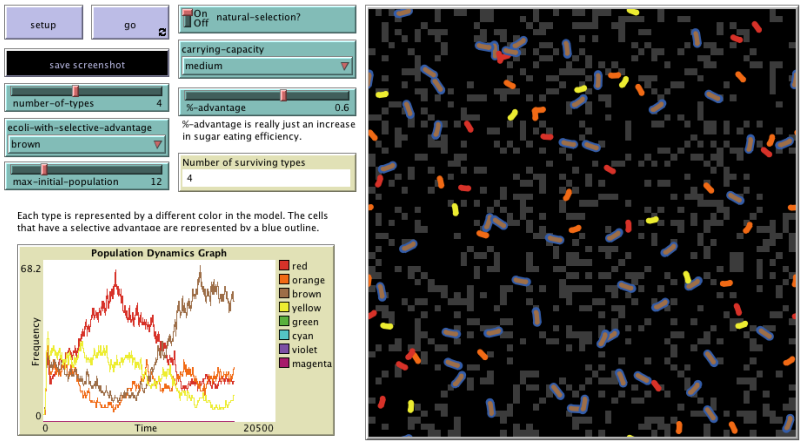


Figure 2-32 Genetic Drift and Natural Selection in NetLogo

2.10.2 CompuCell3D

CompuCell3D (Cickovski et al. 2005) (http://www.compucell3d.org) was designed to study morphogenesis, differentiation and influence of molecular gradients on cell behavior (Figure 2-33). It implements various models discussed earlier like cellular automata for cell localization, agent-based model for individual cell behavior description or rule-based models of molecular diffusion, cell differentiation and evolution. Its plugin architecture allows the addition of more specific types of models.

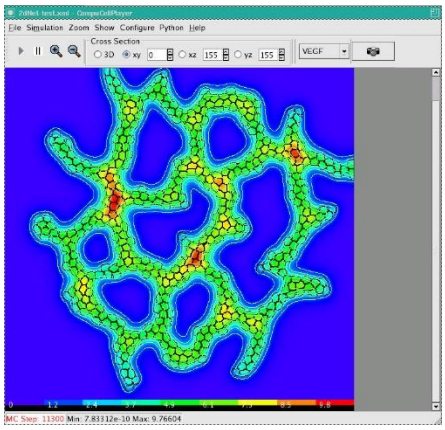


Figure 2-33 Visual output of CompuCell3D

Although the granularity used in CompuCell3D is at the right level for the simulation of communicating cells, the control of cellular level behavior for the emergence of communication is not a straightforward process.

2.10.3 The Virtual Cell

Virtual Cell (Moraru et al. 2008) (http://vcell.org) is based on particle models (see section 2.9.3.3) and includes many algorithms to solve the equations associated with this approach. It also includes rule-based models to simulate important processes that do not need to be observed at the particle level. It is well suited to simulate stochastic events and non-isotropic distribution of molecules (Figure 2-34). Nevertheless, the heavy computation cost involved by such fine-grained simulations renders it inadequate for multi-cellular systems.

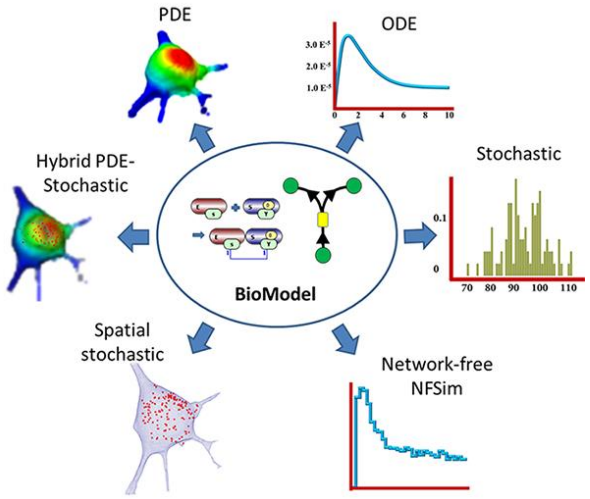


Figure 2-34 VCell applications

2.10.4 CellSys

CellSys (Hoehme and Drasdo 2010) (http://msysbio.com/software/cellsys) studies growth and organization processes in multicellular systems. It uses an agent-based model of elastic and adhesive cells. Cell migration is modeled by an equation of motion for each cell. Cell phenotype can evolve with time through mutations (Figure 2-35).

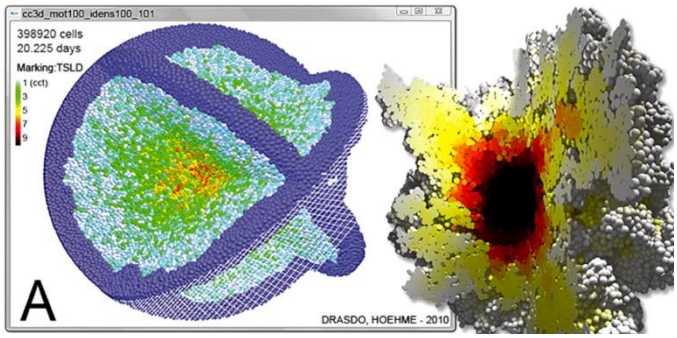


Figure 2-35 CellSys 3D multicellular simulation display

Although this software is designed for multicellular systems it does not include models of internal cell processes nor concentration of molecules in the environment. Furthermore, there is no concept of cellular exchanges between the cells. Its scope is more the simulation of cell aggregate dynamic and the influence of parameters like membrane properties or proliferation speed.

2.10.5 Cell-based Chaste

Cell-based Chaste (Mirams et al. 2013) (http://www.cs.ox.ac.uk/chaste/cell_based_index.html) is a framework initially designed for cardiac electrophysiology and cancer development. Nonetheless recent developments transformed it into a multipurpose tool for multicellular simulation software. It is composed of three modules: 1) A model of cellular behavior, 2) a model of the movement and mechanical interaction between cells and 3) a model of the transport of key nutrients, signaling molecules or waste products. Cellular behavior can range from pure rule-based models to nonlinear differential equations models. At the cell level, it supports both grid-based and lattice-free models. Transport of molecules between cells is modeled using a continuum approach solved using partial differential equation models. An interesting addition is the ability to define various cell killer objects to impose constraints on the cell population and orient mutations (Figure 2-36).

Cell-based Chaste is a C++ library. Although the possibility to set aside mechanical cell-cell interactions makes it an interesting framework for our purpose, the actual implementation of its algorithms in a program is not straightforward and limits its usability.

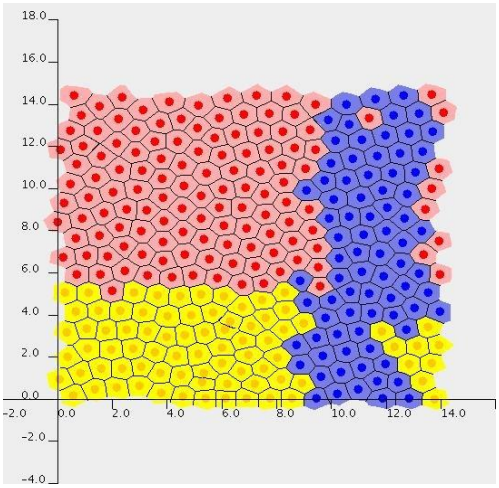


Figure 2-36 Visualization of Cell based Chaste simulation with various mutants

2.10.6 EPISIM

EPISIM (Sutterlin et al. 2013) (http://tigacenter.bioquant.uni-heidelberg.de/episim.html) is a multiscale simulation software for multicellular systems. Process diagrams of cellular behaviors are drawn and then translated into code for the simulation. These diagrams can represent either deterministic and/or stochastic models (Figure 2-37). The simulator code is able to integrate various simulation algorithms like differential equations for diffusion, mechanistic simulation of cell to cell contacts and rule-based models. The easy to learn and use interface allows to quickly build a model for the multicellular simulation. Nevertheless, it can be quite difficult to incorporate custom evolution rules or uncanonical behaviors which makes this software unsuitable for our work.

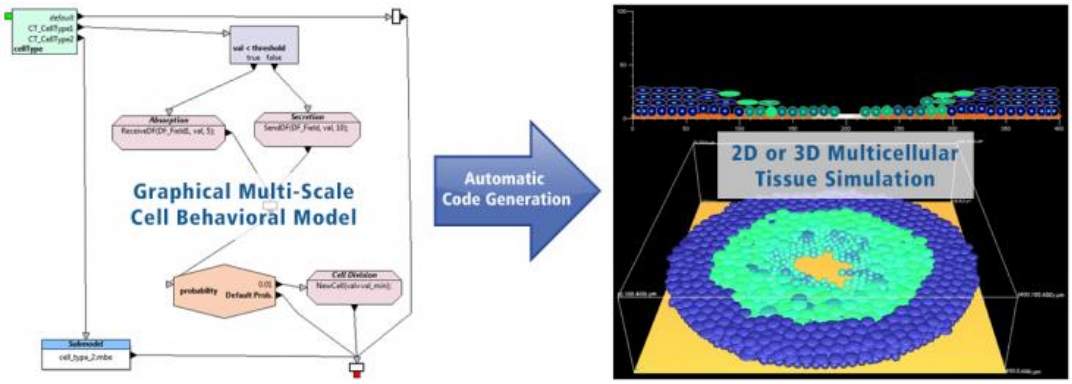


Figure 2-37 Generation of a cell behavior model and then simulation in EPISIM

2.10.7 Morpheus

As for EPISIM, Morpheus (Starruß et al. 2014) intends to render multicellular simulation easy to setup for non-developers (https://imc.zih.tu-dresden.de/wiki/morpheus/doku.php). It is a modeling tool that implements several algorithms like Ordinary Differential Equations (see section 2.9.3.5), Reaction-Diffusion systems and Cellular Potts models (see section 2.9.3.4), and mix them for multicellular, multiscale simulations (Figure 2-38). New projects are relatively quick to setup and the expected parametrization of the various modules is present but adding new custom plugins is not straightforward and can be difficult to interface with other modules in a multi-scale fashion.

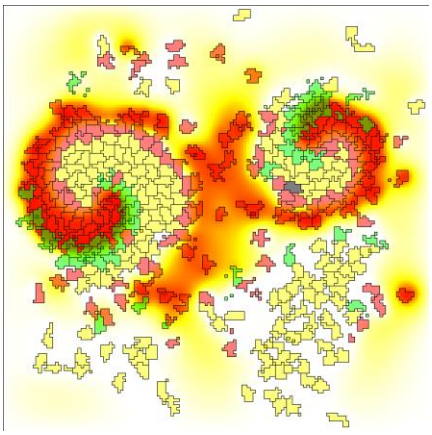


Figure 2-38 Aggregation of amoebas through chemotaxis towards waves of cyclic AMP in Morpheus

2.10.8 LBIBCell

The LBIBCell framework (Tanaka, Sichau, and Iber 2015) implements several useful algorithms for cellular simulation in a 2D medium (https://tanakas.bitbucket.io/lbibcell/index.html). It is not a piece of software *per se* but a library to use during the development of a C++ application. The environment is continuous and cells are considered as elastic polygons that can move and grow. The Lattice Boltzmann method is used to solve diffusion of signals. To solve the interactions between the viscous medium and the elastic cells, the IBcell model is used (Rejniak 2007). Cell behavior can be customized using a plugin architecture. The purpose of this framework is mainly to study morphogenetic phenomena in a multicellular setup (Figure 2-39). This is why the cell-cell contacts, geometry and viscosity of the medium get so much attention. This approach is usually faster than a particle model software.

Although paracrine signals could be simulated accurately, the heavy focus on cell morphology with the associated computational cost renders this library cumbersome for our purpose.

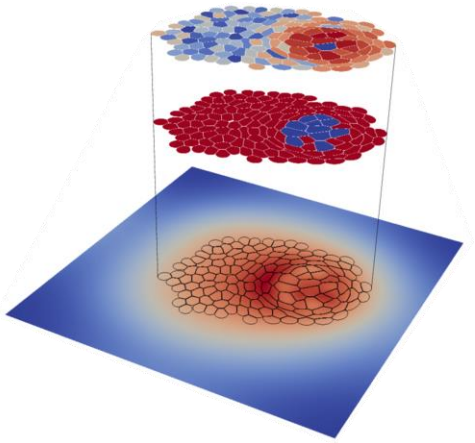


Figure 2-39 Division/differentiation example using LBIBCell. Image generated using Paraview (https://www.paraview.org)

2.10.9 Onko3D

The Onko3D project aims at simulating multicellular tissues within the context of spheroid[7] formation (http://www.itav.fr/portfolio_page/onko3d). Its main focus is cancerous cell proliferation. This simulator is a three-staged one: 1) A behavioral engine that simulates the behavior of cells, 2) a simplified biophysics engine which simulates the physical interactions between cells and 3) a simplified hydrodynamic model that aims at simulating the diffusion of molecular components within the environment.

Our study does not intend to focus on proliferation of cells or cell geometry description, therefore this simulator is not very well suited for our purpose. Furthermore, the project is ongoing and is not at this time available for modifications to address other problems.

2.10.10 NetBioDyn

NetBioDyn (Ballet et al. 2017; Rivière, Ballet, and Rodin 2016) (http://virtulab.univ-brest.fr/netbiodyn.html) is an environment with a graphical interface dedicated to the simulation of multicellular systems and specially designed to be usable by non-developers (Figure 2-40). A multi-agent system is used to describe the cells present in the system, with their behaviors and interactions and also the environment.

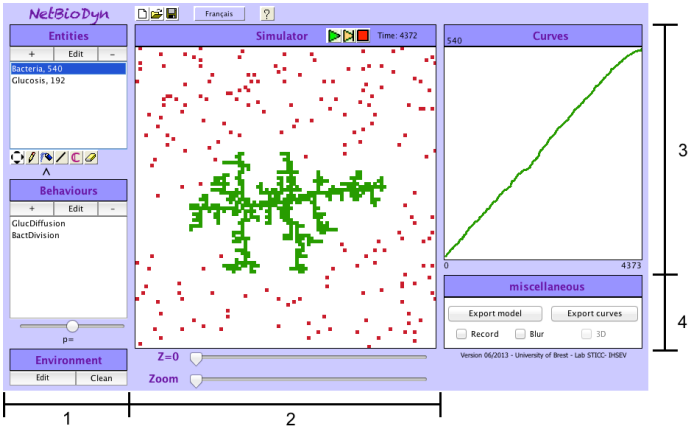


Figure 2-40 Example of bacterial growth in NetBioDyn

One example of experiment describing the use of NetBiodyn deals with a system of two interdependent cell types that reach a dynamic equilibrium by exchanging resources. This is the kind of experiment that typically would need to be performed to study the emergence of communication.

Nevertheless, it is not straightforward to evaluate if it is possible to develop the necessary ingredients for the emergence of communication.

2.10.11 PhysiCell

PhysiCell (Ghaffarizadeh et al. 2018) is a recent multicellular simulator which focuses on simulation of mutual cellular influence through microenvironment and mechanical interactions. It is an agent-based simulation including sub-models for many common cellular events like division, apoptosis or motility. Large systems can be handled (over 500,000 cells) in reasonable simulation

[7] Spheroids are cell cultures in 3D. Unlike 2D culture on classic plastic plates, spheroids are more representatives of *in vivo* cells in the way they behave.

times (Figure 2-41). PhysiCell was originally developed for cancer dynamic studies but it is also suited in other areas of research. It uses a lattice-free, physics-based approach and custom rules can be added to modulate the cell agent behaviors. There are three different time scale used in this framework to account for different mechanisms like molecular diffusion, cell processes and cell mechanics. ODE and PDE are used for the evolution of molecule concentrations in the environment and stochastic events can be taken into account.

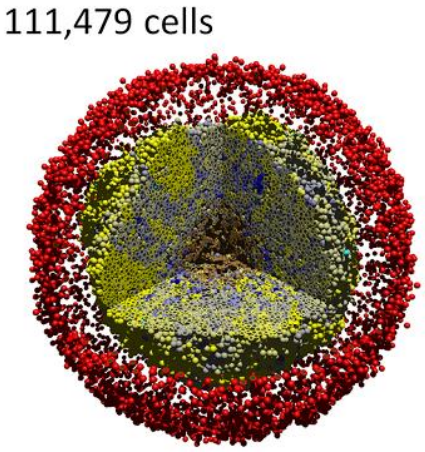


Figure 2-41 3D example in cancer immunology

PhysiCell is currently provided as a C++ library to develop custom problem applications. It contains most if not all necessary models to develop a simulation of communication emergence. Unfortunately, its recent development and availability did not allow us to test its capabilities for applying it to this work. However, it could be beneficial to implement our simulation using this framework to include more up to date ODE/PDE algorithms and distributed computing.

2.10.12 Conclusion on the in Silico Approaches

All the *in silico* approaches mentioned before have been extensively used to simulate various aspects of biological problems. They are often mixed together one way or another to tackle difficult situations where various levels of organization are considered. This leads to specialized software that are often designed for a very specific type of problem. It is extremely difficult to design an all-purpose application that could be used for the vast majority of questions raised by biological experiments. Actually this is not only the case for biology as in other scientific domains the same kinds of problems arise. The main reason for this situation is that we do not know all the rules governing the events happening in the physical phenomena we try to simulate. Furthermore, even if we did, the computational cost of such simulations is usually far beyond the current technological limitations.

When there is no one-size-fits-all solution it is necessary to design the best approach possible for the problem at hand making judicious choices to balance computational requirements and relevance of the model. In the case of the structured communication study, several choices can be readily made:

- Atomic-level or particle-based models are neither tractable nor necessary: A multicellular simulation would require billions of atoms/particles to be in the simulation.
- Since cell communication goes through molecules diffusing in a medium, localization in space is necessary: The model should be grid-based for the cells to have defined neighbors and could use PDE for signal and resource diffusion.

- Each cell is autonomous, can gather information on its surroundings, can take decisions and actions: Agent-based models are well suited for these requirements since their definition and design are very close to those of cells.
- The shape and dynamic of cell contacts might or might not have an impact on the communication: Cell Potts Models might be useful but are not the core of the study.
- Cells will have different capabilities and can evolve with time: A genetic algorithm can be used if a global fitness function can be derived for the system. But in our case the fitness function is difficult to formalize since it is difficult to characterize a communicating system versus a non-communicating one. The other problem of genetic algorithms is that evolution relies on random changes and this can take a very long time to converge.
- Emergence of the communication skills is probably a long process: An algorithm to efficiently explore the parameter space is required without introducing a bias towards the actual appearance of communication.
- Since we are interested in an emergent phenomenon, the simulation must be built from a bottom-up perspective. Multi-agent systems are often used in this situation.

2.11 Emergence of Communication in Simulations

Since the advent of computer-simulated environments, scientists have been interested in population dynamics. Many aspects have been studied like crowd behavior, population migration, social mixing, opinion propagation as well as prey/predator interactions, cancer development, morphogenesis and so on. One recurring field of study in population simulations is communication effectiveness, impact and emergence. Since the 90s many hypotheses have been proposed, systems designed and experiments performed to explain the emergence of communication in communities of entities. What is striking about this particular literature is the fact that no definite answer can be presented as a consensus. Interdependence between the entities and high relatedness seem to be the only two key features commonly found in these studies. In this section are presented some studies dealing with the emergence of communication. It appears that experimental procedures and assumptions are very diverse and most often than not the results and conclusions are linked to these hypotheses.

2.11.1 Communication and Robots

The following examples do not represent an exhaustive sample of the approaches tested in robotic domain. They merely attempt to show some of the results that can be obtained in this field of investigation. One striking aspect of the published experiments is that most of them use the combination of neural networks and genetic algorithms to observe the emergence of communication. Other methods and algorithms exist for machine learning but it looks like this field is focused on this specific combination.

2.11.1.1 Evolution of Symbolic Communication in a Community of Robots

In (Grouchy et al. 2016), the authors start from a non-communicating population of robots. Their setup consists of a toroid world where robots can freely move. One fundamental assumption of their work is the existence of a dedicated channel for communication, namely a sound emitter/receiver. The behavior of a robot is governed by a set of rules that enable movement and communication output. This set of rules represents the genome of the robot and is submitted to genetic evolution in terms of mutations, recombination and offspring inheritance. When two robots

are in close proximity, they reproduce and when a new robot is born (inheriting a mix of its parent's genomes) a random robot of the simulation is killed (Figure 2-42). This process is partly inspired from animal biology since it assumes the existence of sexual reproduction which is not actually the preferred method of proliferation on Earth. In addition, the fact that the birth of a new robot involves the death of another random one is comprehensible in terms of computation cost and for population control but is difficult to justify from a biological standpoint.

An interesting feature of their model is that there is no fitness function involved at any stage of the simulation. The only driving forces at play are survival and reproduction.

Sound emission from a robot is omnidirectional and its intensity is not function of distance. Each robot can only listen to its closest neighbor emitters.

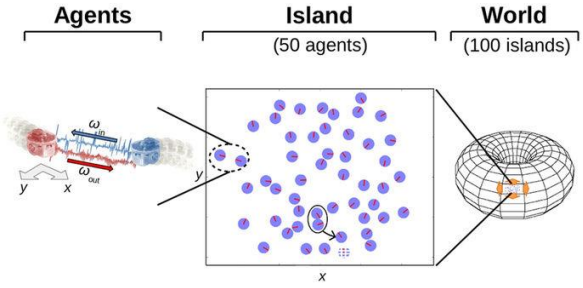


Figure 2-42 General setup of NoiseWorld (Grouchy et al. 2016)

Given this setup and assumptions, the authors were able to demonstrate the emergence of a basic communication protocol between the robots. This protocol was sufficient for an emitter to send information about its location and thus increase the chance of mating with the receiver.

The interesting aspect in this work is the fact that no learning algorithm, substantial preexisting cognitive complexity or explicit fitness function were necessary to observe the emergence of a simple symbolic communication. Still, several assumptions of the simulation limit its genericity and usability in another context. For example, since the evolution of the system is not driven the emergent phenomenon appears after several million simulation cycles. For systems where the agent behavior is slightly time consuming this quickly becomes a hurdle. Also, for the case of cellular evolution, the existence of a dedicated communication channel is a very strong axiom that we would like to avoid.

2.11.1.2 Relatedness and Communication in Evolving Robots

These studies explore the influence of kinship on the communication effectiveness (Floreano et al. 2007; Mitri, Floreano, and Keller 2011). The setup considers a population containing 100 groups of eight foraging robots each with various levels of relatedness. In the environment, exist food sources and poison sources that can only be distinguished at close range (Figure 2-43). As in 2.11.1.1, a dedicated communication channel (in this case light emission/reception) preexists in the robots. This channel can be used to alert other robots of the position of poison or signal the position of food. A fitness function for the robots is evaluated as the time spent around food sources over the time spent around poison. The evolution of the robots is based on mutations and sexual reproduction and is driven by the fitness function. The genome (264 bits) of each robot is the encoding of the small neural network that determines its behaviors. The relatedness between two robots is then defined as the similarity between their genomes.

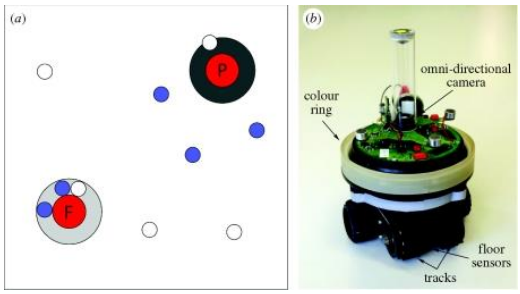


Figure 2-43 (a) Food and poison sources, both emitting red light are placed 1 m from one of two opposite corners of the square arena. Robots (small circles) can distinguish the two sources by sensing the color of the circles of paper placed under each source using their floor sensors when driving over the paper. (b) The robot used for the experiments (Mitri, Floreano, and Keller 2011)

The findings of this study are as follows: Over 500 generations simulations, the overall performance of all robots increased. Group performance is maximal when genome similarity/relatedness is the highest between the robot members. But by performing a test with highly related blind robots (insensitive to communication), the authors show that it is the emergent communication system that is mainly responsible to the high fitness of this group. Although communication is costly for the individual and reduces its fitness, in highly related groups it benefits the group fitness versus competitor groups and represents an advantage on the long run for resource foraging and survival. High relatedness ensures that signals exchanged between robots of a group are reliable, that is, they are sent at the right moment and in the right circumstance by the emitter and properly understood by the receiver.

The results of this study are twofold. First, a useful communication protocol emerges from an evolutionary set of robots. Although a fitness function is used to direct the evolution, it appears quite "natural" in its expression and unbiased towards communication. Second, relatedness correlates well with the reliability of communication and group fitness. This is also observed in real-life organisms and tend to prove that it is a goal to reach to observe the emergence of communication in a competitive environment.

As for the implications for our study they are somehow limited. The neural network that implements the robot behavior is not well-suited for the simulation of cells since it hides too much of the mechanisms taking place during evolution. The genetic algorithm used for the evolution part of the robot system is dependent on the fitness function used. It would be difficult to use it in our case since it is uneasy to define an unbiased fitness function that would distinguish between a system with a communication protocol and without. A closely related work with similar conclusions was described in (Lipson 2007).

Another work using light channel for communication is presented in (Blythe and Scott-Phillips 2014). In this paper, the authors argue that apart from (Quinn 2001) no other published experiments start without preexisting assumptions on the nature of communication. In their opinion, experimental setups always include either: The communication channel, the roles of emitter and receiver, or the forms that signals and/or responses can take. Thus they propose a generic system with no communication assumptions. Although they actually define light channel for communication as in other previous papers, their conclusions are that there are two processes by which communication can emerge from natural selection: Pre-existing actions or pre-existing reactions. That is, some pre-existing actions exist in the agent behavior that could trigger a light activation and benefit to the global fitness of the group, or a pre-existing beneficial reaction that could be triggered by a light signal.

Although using very different setups, these studies reach quite similar conclusions and observe the emergence of basic communication that benefits the community of agents/robots. The fact that the benefit is not to the individuals but to the collective is an interesting finding.

2.11.1.3 Evolving Communication without Dedicated Communication Channels

In (Quinn 2001; Quinn et al. 2003), robots are controlled by neural networks to evolve communication abilities without the existence of a dedicated channel, as sound or light, like in previous experiments. Robots in this setup are only able to control their movements. Their task is to move by pairs a certain distance away from their starting location but staying together without colliding. They are said to stay together if within sensor range. The neural networks are then evolved using a genetic algorithm with a fitness function based on the distance travelled and pairing success. After a few thousand generations, evolved neural networks were able to successfully fulfill the given task. In each case, one of the robots acted as a leader and the other one as a follower. A form of "body language" had emerged. The first robot to move away from the other became the leader and the second assumed the role of follower. The signal being the first movement.

This nice example of communication emergence is particularly interesting since signals appear from a non-signal specific channel that is the movement actuators.

Transposition of this methodology to our purpose is not straightforward. The genericity of this system where there is not assigned communication channel matches well a cellular environment where all resources are equivalent. But as in the previous example, the evolution of a neural network using a genetic algorithm and a fitness function cannot be easily used in our setup and would not be desirable since it could bias the system towards what we want to observe.

2.11.1.4 Indirect Communication in a Competitive Environment

The case described in (McPartland, Nolfi, and Abbass 2005) is interesting because it is closer to a cellular system where cells are in contact only through the environment. Indeed, as for real cells the communication channel is indirect in the sense that robots can only act on the environment by depositing a resource that will be interpreted by other robots as communication. This is very close to an ant simulation where every ant can deposit some pheromones in the environment to indicate its whereabouts to other individuals of the colony. In the paper, two robot teams compete to explore as much territory as possible during a run. Robots are controlled by a neural network optimized using a modified genetic algorithm based on a Pareto multi-fitness approach. One team is able to use indirect communication whereas the second team is more like a collection of individuals. There is competition between the two teams since once a location is explored by a robot (giving one point to the team) it cannot be discovered by the second team anymore. In order to investigate the influence of communication on performance, the authors compared results when communication is off, random and on. When "pheromones" are off, both teams work as individual units. A random use of the communication channel introduces a lot of noise in the system and the performance of the team actually decreases. But when the communication channel is evolved at the same time as the rest of the neural network, the team becomes victorious in all simulations.

The evolution mechanism was able to optimize the use of indirect communication into a competitive advantage. Again this study uses neural networks and genetic algorithms which are approaches we think are best to avoid for our study.

2.11.1.5 Conclusion on Communication of Robots

As shown in the previous examples, simple communication can emerge between independent adapting entities. This phenomenon is easier to observed when a preexisting channel for communication is included in the design of the robotic behavior but still can happen through channels designed for a different purpose (like movements used as body language).

In most cases, a learning algorithm is used to enable robots to adjust their behavior and an evolution process to optimize their population towards a desired global system goal. The most popular learning algorithm is the neural network. It is flexible and able to model very different experimental setups without much change in its structure.

Unfortunately, neural networks are difficult to analyze and even if they correctly do what they are designed for, it is usually hard to understand how and why. Moreover, it may be difficult to justify their use as a model for the internal structure of a cell since it is very far from its actual inner workings (chemical reactions, gene activations…). Furthermore, if communication is observed with a NN, these finding could not be generalized to cells since they do not function this way.

The evolution algorithm used in most experiments derives from a genetic algorithm involving reproduction, random mutations and recombination. This process requires to evaluate a fitness function describing the overall goal of either the individual agents or the whole system. This fitness function can be difficult to formulate/evaluate when the goal of the system is to reach a dynamical equilibrium. Also, as mentioned previously, random mutations are not a computationally effective way of system evolution when simultaneous multiple mutations are required to improve the cell function.

2.11.2 Communication in Biological Simulations

To our knowledge no attempt has been made to model the emergence of communication in the case of multi-cellular organisms. Often, modelling approaches in biology are very focused on a practical problem derived from real-life data and simulation is used to propose new experiments or test hypotheses. More fundamental aspects like the emergence of communication is probably more difficult to deal with since many aspects of the model can be subject to criticism. Indeed, defining the proper model for non-communicating cells is challenging as we do not observe them in nature. Also, the choice of intra-cellular mechanisms to model or not, metabolites and so on are all subject to debate. This is probably why purely biological cellular simulations are hard to find on fundamental topics.

In this section, some examples of what can be found around intercellular communication simulation are given.

2.11.2.1 Cell–Cell Interactions in Regulating Multiple Myeloma Initiating Cell Fate

Peng and colleagues (Tao Peng et al. 2014) developed a novel ODE system to understand how the cell-cell communication regulates multiple myeloma cell fate. The regulation of the four stages of differentiation from myeloma initiating cells (MICs) in primitive progenitor cells (PPCs), committed progenitor cells (CPCs) and mature myeloma cells (MCs) is of particular interest in the study of some cancer. This regulation is heavily controlled by the intercellular communication between these cell types and feedback loops play an important role in the process (Figure 2-44). The secreted factors SSF1, SSF2, ISF1 and ISF2 represent general relationships described in

literature and not specific molecules. This is mainly due to the fact that most of the actual chemical factors are still unknown.

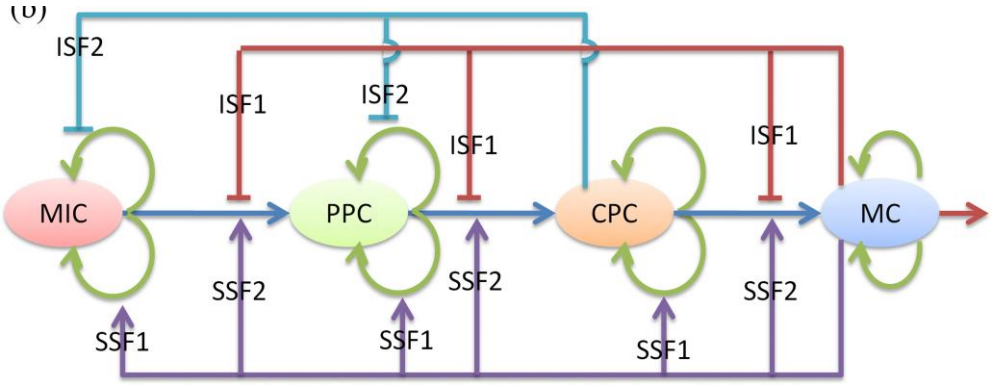


Figure 2-44 Differentiation model of MIC cells (Tao Peng et al. 2014)

A set of 10 differential equations and 21 parameters can represent the system. Experimental data from actual *in vitro* measures performed by the team are used to optimize the values of the 21 parameters. This is a key step in the process of the simulation since the capacity to reproduce real world dynamics is important to enable the use of simulation as a predictive tool.

Although the model was able to reproduce some *in vitro* data, the small set of communication factors and the lack of intra-cellular processes controlling cell behavior and fate limited the use of the model in conditions very different from the data training set.

The hypothesis behind the use of the ODE system is that no perturbation will occur during the simulation time. Also the lack of cell localization supposes that the signaling molecule flow is not hindered by cell division and the concentrations are uniform in the system.

This type of model cannot be used efficiently when simulating the emergence of communication since we do not have a precise knowledge of the way cells worked before communication. Thus, it is difficult to determine a set of equations to govern the evolution of the system.

Another example of ODE model used in cell-cell communication modeling is found in (Shao et al. 2013). The authors modeled the interactions/communication between the immune system and the central nervous system in the amyotrophic lateral sclerosis pathology. The model included 20 differential equations and 70 parameters. As in the previous case, the equations were derived from extensive knowledge of the process under study and literature evidences. The parameters were either found in the literature or optimized to fit experimental data of the system behavior. Once validated, the model could be used to predict the best strategies in order to modulate the behavior of the system out of the pathological state. In particular, the use of paracrine signaling molecule IL6 was counter intuitively predicted to enhance the neuron protecting effect of IL4 better than the direct injection of IL4 itself.

2.11.2.2 Intercellular Network Structure in the Human Hematopoietic System

This study (Qiao et al. 2014) shows that in order to yield interesting results, a static view/data analysis of communication networks can be enough. Ligand production and receptor presences were collected for 12 cell types involved in the hematopoietic stem cell (HSC) system. This consisted in 933 ligand-receptor interactions constructed from gene expression data. Graph analysis and other statistical tools were efficient in order to unravel several mechanisms governing the hematopoietic cell fate. Once several signaling molecules were identified to have major impact on cell fate using the model (Figure 2-45), they were evaluated *in vitro.* 27 of the 33 candidates

proposed by the model showed a major influence on cell fate (quiescence, self-renewal, enhanced proliferation and reduced proliferation). An interesting hypothesis of this study is that in order to use network analysis and propose prediction, the effect of the 33 proposed influential ligands was supposed to be independent of the others. A limitation of the model is that it is static and cannot include dynamic evolution of cellular signaling during the life of the system. Nevertheless, this clearly demonstrates that static data on cell population averages is still information rich.

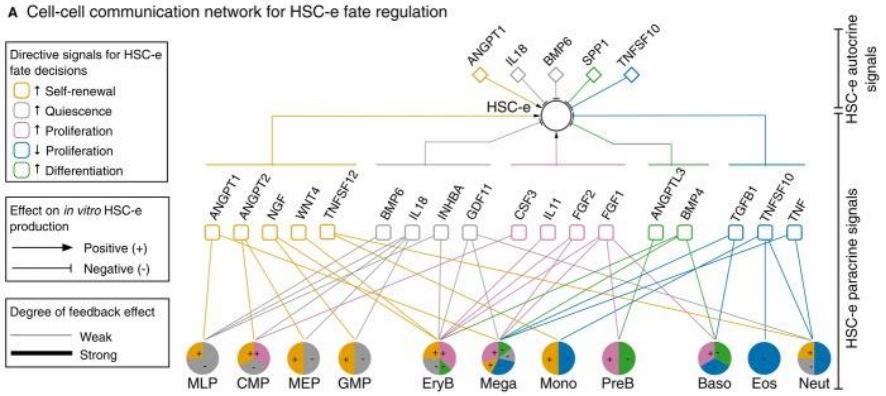


Figure 2-45 Influential signaling molecules regulating HSC fate (Qiao et al. 2014)

2.11.2.3 Cell Communication Networks using Response-Time Distributions

A model of cell-cell communication for simulation purpose is proposed by (Thurley, Wu, and Altschuler 2018). Its purpose is to study the dynamics of extra-cellular communication network using real-life data without the added difficulty of modelling the internal complex machinery of the cell. In order to do so, the cell is considered a black box with time-delayed response to extra-cellular input signals. Temporal input-to-output relationships of intracellular signaling networks are captured by "response-time distributions" (Figure 2-46). These time distributions are not too difficult to measure experimentally and can enrich the model. Most of the time, a gamma distribution can be used to model experimental data. Without entering into too much details about the equations of their simulation, the authors were able to analyze simple network patterns and revealed dynamical properties such as bimodal arrival times and enhanced synchronization, which are masked when treating cell-state changes as molecular reactions. This approach also allowed to interpret actual cytokine secretion patterns.

While this study does not focus on the emergence of communication, it is still interesting in the way it manages cell-cell messages. There is no simulation space and no cells in the model; on the contrary, every aspect of communication is modeled by differential equations that evolve with time and influence the global behavior of the system. The interest of this model lies in the fact that it can be used with real data that can be measured quite easily: For each cellular response, the only needed parameter is the gamma distribution factor.

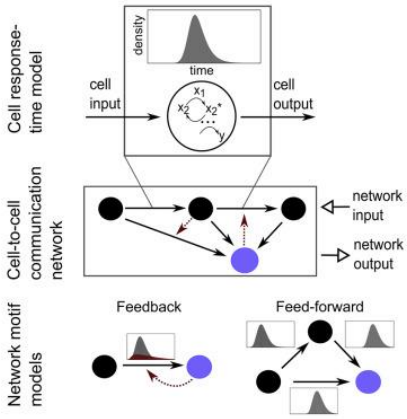


Figure 2-46 Response-time model in (Thurley, Wu, and Altschuler 2018)

There is no direct use of this strategy for our study since we are interested in the emergence of a functional communication system. Nevertheless, gamma distribution could be used as a model for our cell response if we needed to get closer to actual cells and measurable experimental data.

2.11.2.4 Conclusion on Communication in Biological Simulations

Most if not all of the examples about communication in biological simulations that can be found in the literature are tied to real-life applications. The choice of the model depends on the problem to be investigated and on the quality of the experimental data at hand. Parameters in the model are always refined in order to reproduce observed behaviors of the system before it is used for prediction purpose. All in all, this is a very generic use case for simulations: Use knowledge to build a model, adjust it until it fits data, perform predictions to test *in vivo* then start a new loop.

More fundamental work is more the domain of informatics or robotic than biology. This is probably because it is difficult to design a "biological" system that is generic enough to answer essential questions about its behavior and at the same time can pass for actual cells without including data specific to a given biological situation.

2.12 Multi-Agent Systems

As we have seen so far, there are numerous models and software to simulate biological systems, at different scale levels and complexity. In most cases, fidelity of the model to the physical phenomenon is paramount. The direct consequence is that very often, models are complex, precise and costly in terms of computing power. Also, for biological questions that can benefit from simulations, the background is usually well known and experimental data are readily available to direct the design of the simulation. In some cases, the experimental data are used as a target for the optimization of the simulation's various parameters. Knowing the end result of the simulation is a very powerful way to drive it using fitness functions dependent models.

When it comes to more fundamental problems, like evolution of life on Earth, multicellular development, cell communication or other similar themes, there is no specific or generic tool to address them. In these cases, it is often required to develop methodologies and tool adapted to the problem at hand given the specific simplifications and constraints. As discussed, for cell communication, we need to put ourselves in a context where cells are early versions of what is observed nowadays. There are not much data available about this time (unless we consider their descendants as faithful representatives), so it is reasonable to start by using a generic system that represents an abstraction of a real cell of those ancient times. Most of the various algorithms and

models discussed earlier require actual data to be efficient. Since we do not have access to such data, we can only use generic methods that manipulate large sets of individual entities.

Usually computational approaches used in actual published papers about communication have a common characteristic: They require some sort of evaluation of how well the system behaves as a feedback information to direct its evolution. If we use such a fitness function in our system, two problem arise: First it might be difficult to formalize a function that would differentiate efficiently between a system where cells communicate together from a system without communication. Secondly, if such a function can be written and evaluated, the system would probably evolve a communication protocol, but what would be the conditions for its emergence? Since the fitness function directs the evolution of the various parameters in the cell behavior, it becomes very difficult to point out factors that favor the emergence of communication. These two problems are major hurdles and we decided to rule out any methodology that would require such global fitness functions like genetic algorithms or neural networks.

Finally, we need a method that allows to efficiently explore the parameter space of the system to minimize the simulation time. Random walk and stochastic methods are not possible given the time and processing power at our disposal.

As described in the next paragraphs, Multi-Agent Systems (and their extension AMAS) are methods that do not require fitness functions and therefore can avoid random walk in parameter space. Thus, selecting this approach to address our problem seems a sound choice.

According to (Ferber 1999), a multi-agent system (MAS) is composed of autonomous entities, the agents, and the environment in which these agents live. These concepts are discussed in the following paragraphs.

2.12.1 Agent

There are various definitions of the term "agent". A commonly accepted definition was given by Weiss as "an agent is a computer system that is situated in some environment, and that is capable of autonomous actions in this environment in order to meet its design objectives" (Weiss 1999). Ferber added the notion of locality to extend the notion of agent as "an autonomous physical or virtual entity able to act (or communicate) in a given environment given local perceptions and partial knowledge. An agent acts in order to reach a local objective given its local competence" (Ferber 1999).

An agent possesses the following fundamental features:

- It is autonomous, which means that it controls its own behavior. The choice to act or not is only driven by the agent's own behavior;
- It possesses its own resources and skills;
- It evolves in an environment in which it is able to locally act;
- It possesses a partial knowledge of this environment;
- It is able to interact and communicate with other agents either directly or through the environment;

The agent's behavior is based on a Perception-Decision-Action cycle involving three phases (Figure 2-47):

- The perception phase during which the agent gathers information from its environment.
- The decision phase during which the agent decides of the actions to perform. This decision is based on its local perceptions, its internal knowledge and its own objectives.

- The action phase during which the agent performs the previously selected actions.

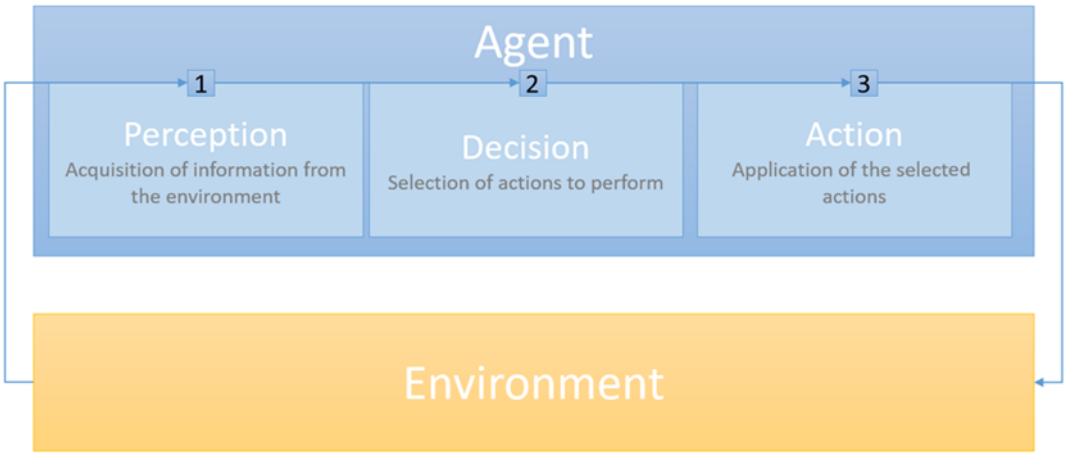


Figure 2-47 Lifecycle of an agent

Agents also possess characteristics that enable their categorization as follows (Di Marzo Serugendo, Gleizes, and Karageorgos 2011):

- Proactive agents versus Reactive agents: A reactive agent's behavior is triggered by events in its environment. The trigger will also depend on the internal state of the agent. This kind of agent does not possess a lot of memory if any. Its actions are reflex based. A proactive agent on the opposite is able to modify its objectives and create new ones. It is also referred as cognitive agent since it can learn from its actions. These two types of agents are the extremes of a continuum of agent configurations. When a system contains a huge number of simple reactive agents it is said to have a fine-grained granularity. If it contains less agents but with more processing power, it is said to have a coarse-grained granularity.
- Situated agents versus Social agents: An agent is said situated if its perceptions and its communications skills are conditioned by its location in the environment. On the opposite, the agent is said social if its perceptions and communications skills are not dependent on its localization. However, agents may directly interact without requiring localization condition. Here too, there is no concrete frontier between situated and social agents.

Those characteristics are not exclusive but they illustrate the expressiveness of the agent-based modelling.

A key component of agent-based modelling is the environment, which will be discussed in the next section.

2.12.2 The Environment

The notion of environment is central to Multi-Agent Systems. Indeed, the environment is not only a source of information, but also the medium for agent's actions and interactions. Although it is a key component of MAS, the environment is not formally defined and there is no consensus inside the MAS community (Weyns et al. 2005). The environment of an agent can be described as everything which is not this entity. Depending on the adopted context, various environments can be identified. Two levels are relevant: The macro-level (MAS's viewpoint) or the micro-level (agent's viewpoint).

- From the system's point of view, the environment is everything that is outside of the MAS.

- From the agent's point of view, the environment is not only a part of the MAS's environment, but also the other agents. The agent's environment is twofold: The physical part of the agent's environment, describing what the agent can perceive and how it can act, and the social part of the agent's environment, describing with which agents it is able to interact.

(Russell and Norvig 1995) propose four characteristics to describe the environment of an agent:

- Accessible/Inaccessible: The agent's environment is accessible by an agent if the agent is able to perceive all the information required for its task.
- Discrete/Continuous: The agent's environment is discrete if it possesses a finite number of distinct states.
- Deterministic/Non deterministic: The agent's environment is deterministic if its evolution consecutive to an action is only dependent on its current state.
- Dynamic/Static: The agent's environment is dynamic if it evolves by itself.

2.12.3 Properties of Multi-Agent Systems

In Multi-Agent Systems, no agent can solve the global problem since it only has partial information or capabilities. Its viewpoint is usually quite limited. However, all the knowledge and competences required for solving the problem are still present, distributed among the system. Thanks to this distribution, the MAS paradigm seems particularly suited to problems with a natural distribution such as biological systems.

A MAS is said to be open if agents can appear and disappear during the lifetime of the system. Otherwise it is said to be closed. The creation of a new agent results commonly from the decision of an existing agent while its destruction can be initiated by the environment or the agent itself (which then commits a form of suicide).

Another very important property is the absence of an external or global control system. The control is distributed over the agents, and each agent is solely responsible for its own behavior.

Finally, a MAS can be composed of agents with the same capacities (homogeneous MAS) or composed of agents with different skills (heterogeneous MAS).

2.12.4 Self-organization in MAS

Self-organization is a spontaneous process of some systems where positive feedback loop mechanisms will alter the interactions structure and behaviors of the components of the system. This process is not driven by any kind of external control and is distributed over the components of the system. It usually results in a more robust organization and is capable to withstand perturbations from the environment. The concept of self-organization has been studied since the ancient Greek in a variety of domain. (Di Marzo Serugendo, Gleizes, and Karageorgos 2011) propose to define this concept from a software engineering point of view: Self-organization is the process enabling a software to dynamically alter its internal organization (structure and functionality) during its execution time without any explicit external directing mechanism.

Some MAS are intertwined with their environment. The actions of a MAS alter its environment and changes in the environment will in turn influence this MAS's behavior and structure in a feedback loop mechanism. Consequently, MAS can be self-organizing.

A difference is made between weak self-organization, where the control of the inner organization is centralized by an internal entity, and strong self-organization, where this control is decentralized:

- Strong self-organizing systems are defined as systems where self-organization process decision is distributed locally among the system components without involving any centralized point of control (either internal or external).
- Weak self-organizing systems are systems where, from an internal point of view, self-organization is internally administrated by a centralized point of planning and control.

Many mechanisms to enable self-organization can be found in literature. Stigmergy, for example, is a mechanism for indirect coordination between agents through modification of the environment without any centralized control of the self-organization process (Mano et al. 2006). Stigmergy seems to be the natural approach for cellular simulation since cells interact together through the medium. Nevertheless, it is often used when agents already know/have a way to get information or inform other agents of their internal state. While studying the emergence of communication between cells, this new phenomenon must be the result of self-organization and not the other way around.

Another self-organization mechanism is the holonic approach (Calabrese et al. 2010) where the system is built in hierarchical layers of agents where the higher levels of agents exercise a direct control on the sub-layers. The holonic approach is not well suited for cells since there is no apparent hierarchical relationship between them.

From those definitions, we can identify two processes in a MAS: One that operates the self-organization and one that realizes the function for which the system has been designed for. However, these two processes are so intertwined that it is difficult to tell if one interaction belongs to the former or to the latter.

Self-organization enables adaptation in MAS. Any change in the organization involves a change in the global function. Then, self-organization can be compared to a form of learning, as the system learns to interact with its environment. But most of all, self-organization is a key concept to control the emergence of desired properties. In the next section, we survey an approach to design artificial systems with emergent functionalities: The Adaptive Multi-Agent System (AMAS) approach.

2.13 Designing the Emergence: The AMAS Approach

The Adaptive Multi-Agent Systems (AMAS) approach proposes a method to study and build artificial systems with emergent functionalities. The capacity of adaptation of an AMAS comes from its capacity to self-organize thanks to the cooperative behavior of agents composing the system (Georgé, Gleizes, and Glize 2003). The principles of this approach are presented thereafter.

2.13.1 Interaction and Cooperation

In the previous section, we have discussed the coupling between a MAS and its environment. The activity of the MAS influences its environment, and the MAS activity is influenced by its environment. According to (Kalenka and Jennings 1999), three types of interactions may exist between a system and its environment :

- Cooperative interactions: The action of an entity promotes the activity of another one providing both entities with individual benefits.
- Neutral interactions: The action of an entity neither hinders nor promotes the activity of another one.
- Antinomic interactions: The action of an entity hinders the activity of another one.

A system is in a cooperative state if all its interactions are cooperative. A system is in a non-cooperative state if there is at least one interaction that is neutral or antinomic.

2.13.2 Functional Adequacy

An artificial system is designed to perform a function, so it is said to be functionally adequate when it executes the function for which it has been designed. Usually, the evaluation of the functional adequacy is determined by an external entity which observes the system activity. However, with a MAS, this evaluation has to be performed by the inner agents which have no clue on the global task. This must rest on self-observation capacities, evaluating only local criteria. The Adaptive Multi-Agent Systems approach stipulates that a system in which all the agents are in a cooperative state is functionally adequate (Georgé, Gleizes, and Glize 2003). The AMAS approach proposes a definition of the functional adequacy based on the categorization of the interactions between a system and its environment: A system is functionally adequate if it has no antinomic activity on its environment.

Therefore, a cooperative system, which has only beneficial activities with its environment, is functionally adequate.

Given this definition, (Glize 2001) expresses the theorem of functional adequacy as: Given a functionally adequate system, there exists at least one cooperative internal medium system that fulfils an equivalent function in the same environment.

A cooperative internal medium is a system in which all the interactions between its constituting parts are cooperative. For more information on the demonstration of this theorem, the reader can refer to (Georgé, Edmonds, and Glize 2004).

So, for each problem where a solution is actually calculable, there exists a MAS that solves this problem in which all the agents are in a cooperative state. Thus, the design of a functionally adequate system can be made with a focus on the design of local cooperative interactions between its constituting parts.

2.13.3 Adapting the System through its Parts

A (Adaptive) Multi-Agent System is inherently linked with its environment. When there is a change in the environment, the system may not be in functional adequacy anymore. From the functional adequacy theorem, non-adequacy in an AMAS system comes from non-cooperative interactions between agents. In order to recover the system functionality and revert to a functionally adequate state, agents within the system need to locally detect the non-cooperative interactions and change their behavior accordingly. Therefore, self-organization of an AMAS rests on the self-observation capacities of its agents to detect, anticipate, and prevent or solve failures in its perception, decision or action processes resulting in non-cooperative interactions. Seven types of such failures, called "Non Cooperative Situations" (NCS), have been identified, the first two may appear during the Perception phase, the next two may happen during the Decision phase and the last three during the Action phase:

- Incomprehension: The agent cannot extract the semantic contents of a received piece of information
- Ambiguity: The agent extracts several interpretations of the same message.
- Incompetency: The agent is unable to use the available information to take a decision.
- Unproductiveness: The agent cannot propose an action to do.

- Concurrency: The agent perceives another agent which is acting to reach the same world state.
- Conflict: The agent believes that the transformation it is going to operate on the world is incompatible with the activity of another agent.
- Uselessness: The agent believes that its action cannot change the state of the world or it believes that the results for its action are not interesting for other agents.

Consequently, the behavior of a cooperative agent can be split into two parts:

- Nominal behavior which ensures the functional adequacy when the agent is not faced with Non Cooperative Situations.
- Cooperative behavior, a subsumption of the nominal behavior, which enables the agent to reach its nominal behavior by solving NCSs by locally adjusting its behavior using three possible means (Capera 2005):
    - Tuning: The agent adjusts its internal parameters.
    - Reorganization: The agent changes the way it interacts with its neighborhood, i.e. it stops interacting with a given neighbor, or it starts interacting with a new neighbor, or it updates the confidence given to its existing neighbors.
    - Evolution: The agent creates one or several other agents, or removes itself.

The cooperation in an AMAS is provided by mechanisms which either anticipate or resolve NCSs. This task is devolved to the system designer who has to identify the possible NCSs and to propose adequate mechanisms to solve them. In the next section, we present a methodology to build Adaptive Multi-Agent Systems.

2.13.4 The ADELFE Methodology

The AMAS approach differs from traditional MAS engineering by its focus on local cooperative behaviors. The designer must describe the system's environment, specify the agents composing the system, characterize their interactions and failures in cooperation, and propose mechanisms to restore a cooperative state if needed.

ADELFE (Bernon et al. 2003; Picard and Gleizes 2004) is the French acronym for "Atelier de Développement de Logiciels à Fonctionnalité Emergente" which can be translated by Toolkit for Designing Software with Emergent Functionalities. The ADELFE methodology is based on the well-known software development methodology Rational Unified Process in which some workproducts specific to the AMAS approach are added (Bonjean et al. 2014). ADELFE is composed of 21 workproducts split into five main work definitions:

- WD1 - Preliminary requirements: This phase represents a consensus description of specifications between customers, users and designers on what must be and what must do the system, its limitations and constraints.
- WD2 - Final requirements: In this work definition, the system achieved with the preliminary requirements is transformed into a use case model, and the requirements (functional or not) and their priorities are organized and managed.
- WD3 - Analysis: The analysis begins with a study or analysis of the domain. Then, identification and definition of agents are processed. The analysis phase defines an understanding view of the system, its structure in terms of components and identifies whether an AMAS is required to design the system.

- WD4 - Design: This phase details the system architecture in terms of modules, subsystems, objects and agents. These activities are important from a multi-agent point of view since a recursive characterization of multi-agent systems is achieved at this point.
- WD5 - Implementation: In this work definition, the framework and agent behaviors are implemented.

The ADELFE process is not a simple waterfall process as some loops and increments are included. A complete description of the ADELFE approach can be found in (Bonjean et al. 2014).

2.14 Conclusion on Biological and Informatics Background

In this chapter, we have seen that cells in higher organisms live in communities. To coordinate their actions, they exchange information in the form of molecules. These signaling molecules are often believed to work on their own, *i.e*. one signaling molecule induces one behavior. Acknowledging that Nature often optimizes life processes, the structured communication hypothesis of this thesis is that signaling molecules are like words and only combinations of them into sentences carry meaning to the cells receiving it. Published evidence in favor of this hypothesis were presented.

Direct proof of the hypothesis through physical observation would be very difficult. Firstly, because the cellular system must be well known and studied in an environment very close to *in vivo* conditions. Secondly, such experiments would quickly face the combinatorial explosion problem. Thirdly, unexpected cell behaviors induced by signal combinations would be very difficult to detect since they are… unexpected. Computer simulation can provide means to test evolution scenario where communication between cells emerges. If the simulated system is designed in a way that is generic enough to represent a real cellular system but not to precise to avoid the lack of knowledge and data about it, then the findings of such simulations can have interesting reaches and implications.

From the survey of the available methodologies in biological simulations, it appears that either a lot of prior knowledge about the system and its function is needed or a huge amount of computing power is required to bypass the unknown information using low-level physics. No ready-to-use solution exists to address general multicellular systems when very long evolution time is required and the global function is unknown.

Using the AMAS approach, it is possible to design a system with very little prerequisite knowledge of the cell inner workings but at the same time generic enough to represent the cell role in a multicellular organism. Furthermore, cooperation between agents is the core of the AMAS approach and can be used to boost the speed of the exploration of parameter space without introducing any other bias than the cooperation axiom itself.

CHAPTER 3. CONTRIBUTION: COCELL

Pour les raisons présentées dans le chapitre précédent, notre contribution à l'étude de l'émergence de la communication dans les systèmes multicellulaires implique la construction d'un tissu multicellulaire simulé basé sur un système adaptatif multi-agent (AMAS). Pour simplifier la discussion, cet AMAS est appelé CoCell (pour « Communicating Cells »).

L'objectif de ce chapitre est de jeter les bases de notre contribution en motivant les choix stratégiques effectués, en donnant l'architecture générale de CoCell et en détaillant les comportements de ses agents coopératifs.

CoCell est développé en trois étapes incrémentales qui sont détaillées et évaluées dans les trois chapitres suivants : CoCell1 traite de la mise en œuvre des concepts de base, de leur pertinence en termes de stabilité, d'efficacité, d'évolutivité et d'absence de biais, CoCell2 introduit les mutations et leur impact sur la dynamique du système cellulaire et, enfin, CoCell3 se concentre sur l'émergence de la communication entre cellules.

For the reasons presented in the previous chapter, our contribution to the study of communication emergence in multicellular systems involves building a simulated multicellular tissue based on an Adaptive Multi-Agent System (AMAS). To simplify the speech, this AMAS is called CoCell (for Communicating Cells).

The aim of this chapter is to lay the foundations of our contribution by motivating the strategic choices made, by giving the general architecture of CoCell and by detailing the behaviors of its cooperative agents.

CoCell is developed in three incremental steps which are detailed and evaluated in the next three chapters: CoCell1 deals with the implementation of the basic concepts, their relevance in terms of stability, efficiency, scalability and their lack of bias, CoCell2 introduces mutations and their impact on the dynamics of the cellular system, and, finally, CoCell3 focuses on the emergence of communication between cells.

3.1 Strategic Choices

The general problem of the CoCell project is to study by simulation the exchange of materials through the environment that would act as communication signals in order to maximize the survival capacities of a multicellular tissue. In order to avoid any bias in the results, we have defined strict conditions in the simulation that we can summarize in three points:

- A cell does not know what gives itself the ability to survive in its environment (its energy source and the rules to produce or gather it).
- Consistent with biological knowledge, the decision of a simulated cell must be strictly local.
- No comprehensive knowledge of the global state of the system can guide this local decision.

The emergence of the phenomenon that we wish to observe induces us to avoid any presupposition that humans may know about biology: Notably the materials necessary for survival, the rules of transformation essential to the cell, the reserves that it should carry out, the acquisition or release of materials into the environment...

We therefore place this work in a context with much less information than what a computer scientist may want to inject into a simulation to facilitate its convergence towards a "satisfactory" state. The first consequence of observing the emergence of these survival conditions without guiding them is to ban all methods that use cost/evaluation functions to guide cell decision making. Thus, this comforts the conclusion already made in the previous chapter, that supervised methods such as genetic algorithms, neural networks, swarms... are inappropriate for our purpose.

The first step is to define the context of the simulation. The difficulty is to add every relevant aspect of the real world for the topic being investigated without including too much details. The more elaborate the simulation the more difficult it is to interpret its behavior. Indeed, each simulated feature needs a set of specific parameters to be tuned and this quickly becomes intractable. Also, the simulation must be computationally optimal and the cell model generic enough to be representative of its natural counterpart. Both of these features present their specific challenges and are usually mutually exclusive.

We are interested in the emergence of communication in an interdependent set of cells. At the same time, from the informatics side, we are looking at the cooperation and coordination of a system of entities without preexisting communication protocols. To gather interesting information about communication, the model must abstract most of the actual physical processes found in real cells. For example, enzymes are able to catalyze chemical reactions, and they are made of thousands of atoms in a precise three dimensional conformation. For our purpose, we can abstract an enzyme to just how much it accelerates a given reaction without focusing on its physical conformation.

Abstraction brings simplification and genericity but could also introduce biases. Simplification of complex processes is beneficial in terms of computational power. Genericity is advantageous because there is no need for a precise description of a real system to start implementing the simulation. The reaction constants, initial concentrations of molecules and activated genes for an accurate cell simulation are difficult to measure and to be complete. Since this kind of simulation is chaotic in nature, sensitivity to initial conditions is very high and small errors on measured parameters may lead to important errors during simulation. Defining a generic system that does not represent an accurate cell system frees us from this kind of constraints. On the negative side, if the system is too generic there is no easy way to check its behavior against actual physical data. Only self-coherence and our own judgement can assert if the system works properly or not. This can be a very strong bias since nobody is fully objective, and we do not exactly know what to consider as a normal behavior.

Communication exists to share information and to request modification of behavior. To be of any use, communication must be needed. That is, the cells in the system must have different capacities and be able to be of service to one another. A form of interdependency is in order, otherwise cells would behave as most single cell organisms: Self-contained autonomous organisms that consider anything other than themselves as either threat or food.

The first initial step to approach the problem was to create a collection of cells with a working communication system already in place. From this, we tried to observe the conditions for the emergence of a combinatorial communication protocol. But this kind of system required too much tuning to work properly and we were never sure if what we observed was an emergent property of

this type of systems or the result of the bias introduced by all the information put into their construction and tuning.

We concluded that it was necessary to start from a very basic system with simple cells much like unicellular organisms. From this system and using a fast evolving algorithm based on cooperation, cells would generate their own communication system with as little bias as possible.

The adopted approach for this is to build the system step by step, check that its behavior matched our expectations and that no bias is introduced by our methodology. Broadly this work can be split into the following areas:

- Define the environment, the cells and their interactions/interdependencies.
- Define the inner workings of a cell.
- Carry out the cooperation between cells and study how it can speed up the exploration of the parameter space.
- Define cell death and how to replace a dead cell.
- Add cell mutations to help the system converge towards a viable dynamic equilibrium.
- Define and study interdependence scenarios to observe communication emergence.

MAS models are already used in many biological simulations as presented in section 2.9.3.13. To implement a simulation of a multicellular system with all the constraints and choices described earlier we will use the AMAS approach since its related concepts such as self-organization and cooperation enable to design more easily the agent's life cycle that leads to the intended emergent behavior of the system. Furthermore, cooperation also accelerates the exploration of the system parameter space and thus reduces computation requirements.

Before describing the Implementation details, some elements of the system design need to be discussed. First, an important aspect of multicellular organism's life, dynamic equilibrium or homeostasis, is defined, then the shape of the simulation space is questioned since it can influence the system dynamics, and finally, the time scale of the simulation is discussed for the types of phenomena we want to observe.

3.1.1 Dynamic Equilibrium or Homeostasis

Equilibrium is an extremely strong drive in Nature. Most systems are changing only to reach this state. It is usually synonymous to a locally minimal energy state. Once reached and unless an external perturbation occurs, a system will stay in an equilibrated state.

A different form of equilibrium exists: Dynamic equilibrium, or homeostasis for biological systems. As opposed to a stable equilibrium, a dynamic equilibrium is a balanced state which requires negative or positive feedbacks to stay in its steady state (Figure 3-1). In order to remain functional or to operate in an efficient manner, complex dynamical or open systems often need to reach such a dynamic equilibrium. Such systems are found in very diverse areas like economy (Aruoba, Fernandez-Villaverde, and Rubio-Ramirez 2006), finance (Shimomura 1998), ecology (Tuljapurkar and Semura 1977) and very often in biology (Bernado and Blackledge 2010). Cells are example of such systems. They are permanently receiving information, taking actions that modify their internal state and adjusting their behavior. But from the outside they may appear as if they are at equilibrium since their overall state does not change. The same happens for multicellular communities where cells are adjusting all the time to maintain their main function in the collective. From the outside, the system looks as if it is at equilibrium since its resources inputs and outputs are nearly constant.

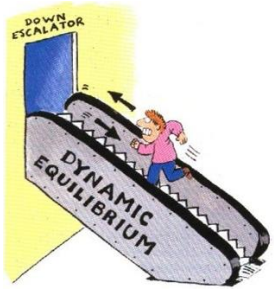


Figure 3-1 An example of dynamic equilibrium

In order to decide if a system is functioning nominally, the criterion used across this study is first the establishment of a dynamic equilibrium. A secondary criterion is the average lifespan of cells for preventing the system to reach a dynamic equilibrium while the cells are being renewed nearly permanently. From the real life examples, this does not appear as a valid solution for a "good" system.

Theoretically, a dynamic equilibrium should be reached for every cell to continue to exist and share resources. In principle, achieving such a state is possible if actions performed by the system's cells are specifically suited to give the feedbacks required to maintain the balance.

3.1.2 System Space

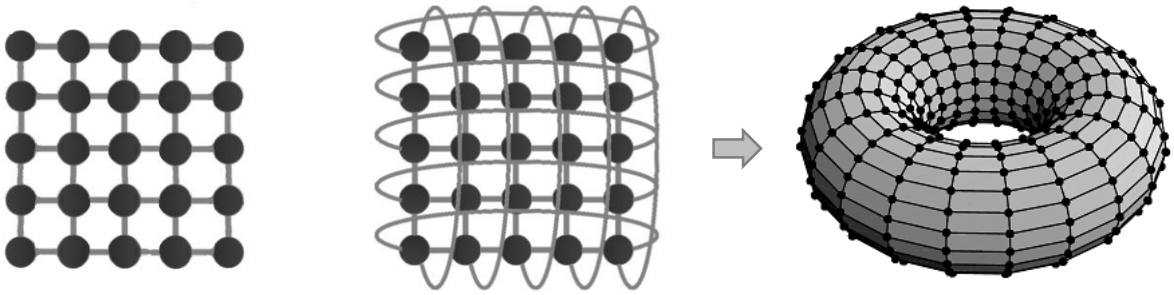


Figure 3-2 System space

An important aspect is the shape of the environment where the simulation takes place. In order for each cell to have a definite number of neighbors and for messages to travel isotropically from their emission point, the natural implementation is a grid. Each node of the grid contains a cell and resources. Thus, in this setup the number of cells is directly linked to the size of the grid. This grid can be either 2D or 3D, the main difference being the number of neighbors for each cell. The edges and corners of a grid are special cases and can be problematic since the number of neighbors is not the same as in other places. To avoid this, the left edge is connected to the right edge and the top edge to the bottom edge. This gives the grid the shape of a torus in a higher dimension (Figure 3-2). The arrangement of the nodes on the grid could have some influence if the distance between nodes is used in some part of the simulation. In 2D the best distribution to keep equal distances is the hexagon. Nevertheless, in our case distances are not used, making a square grid a good representation. This can be easily modified since we only define our system by the neighbors of each node. The coordinates we use are merely for indexing and display purpose.

In an organism, some types of cells are able to move and to change size. For example, white cells can trace pathogens back to their location by following concentration gradients of specific molecules. Adipocytes are cells that store lipids (fat) for future use and can increase their size up to a factor of 50. Although, cell movement and size are important for some aspects of an organism

life, many cell types do not use these capacities much. As such, in CoCell, cells will not move or change size, and will stay on their grid node. The fixed size of cells comes with a drawback though. During cellular division, resources from the mother cell are equally divided between the daughter cells. But the size of the daughter is also half of their mother. So the actual concentration of resources stays the same although their quantity is divided by two. This cannot happen with cells that cannot change size. This problem is discussed further in paragraph 3.2.4.5.

3.1.3 Time Scale

The ambitions of our simulation are in a way contradictory. We aim at creating a simulation where cells will evolve a communication system to improve their survival performance. Evolution in terms of natural organisms is measured in millennia. Survival performance on the other hand is an instantaneous measure and affects every action of the organism. In other words, we are trying in CoCell to merge two very different time scales. This problem is not specific to this particular theme. As was mentioned in section 2.9.3, multi-scale simulations are faced with similar hurdles when mixing atom-based molecular dynamics with cellular-level gradient formation or gene regulation. To effectively mix scales and models there is no absolute right way. It is more a matter of choice and measured impact on the global system behavior.

In CoCell, we choose to focus on the survival performance timescale. That is, given a set of abilities the cells will work together to try to survive by exchanging resources in "real" time. The evolution timescale is somehow hidden in the mutation process that occurs when a cell dies and is replaced. This choice is justified by the fact that when the communication protocol emerge, the system should dramatically improve its survival performance. But this can only be observed during the life of cells, that is, at a fine grain time scale.

Thus, we define a time cycle that is more in phase with resource diffusion and cellular actions. During a time cycle each cell is able to perform an action and resources in the environment can diffuse from one node to another. Even this definition of a time cycle can be seen as arbitrary since a cell is theoretically able to perform multiple actions in parallel. Nevertheless, we consider that most of these actions are part of the background cell management and that only one action per cycle is relevant to our study. To take into account this background work, the energy of each cell is decreased by a fixed amount every cycle.

3.2 Implementation of the Simulation as an AMAS

From the high-level description of the system and phenomenon we want to study, the following structure is derived for the implementation of an AMAS.

The system is composed of cells living in a common medium. Material resources are present inside the cells as well as in the environment where they diffuse freely. The system is defined on a 2D or 3D grid and to ease several aspects of the simulation, cells are considered immobile, of fixed sized and located on the nodes of the grid.

3.2.1 System Agentification

Cells are autonomous, decision making entities that are able to perceive information about themselves and their local environment. This is a perfect match for the classical agent definition. Thus, cells perform the typical agent life cycle, perception-decision-action.

The role of a cell agent is basically to perform resources transformations. If the transformations are adapted to its environment, the cell survives.

Since we want study "early" cells that evolve communication and interdependency, internal regulatory mechanisms are kept to a minimum and the internal structure intentionally kept quite simple. An advanced implementation of a cell could consider it as being an AMAS system itself but we preferred to start with a simple rule-based model and leave an internal AMAS for future developments.

Cell agents are described in more details in section 3.2.5. In our case and since communication between cells is what we want to observe, it cannot be included in the agent description. Therefore, the cell agents as described here lack this feature to perfectly match the classical definition where agents are able to directly communicate by exchanging messages. Consequently, the simulated cells are unable to exchange any information, about their behavior, their internal state or their desires, with their direct or indirect neighbors. Since most of the time, in AMAS, this restriction does not exist, this aspect is very challenging and will enable us to study how agents may still cooperate and make a collective behavior emerge with so little information.

A grid node where a cell agent is located, and although fundamentally different from it, is also an agent since it decides to perform actions based on local perceptions. Its role is to move material resources between cells (diffusion), check if the cell it contains is alive and if not replace it with a more suitable one from its neighborhood. The node agent is described in section 3.2.4.

The system also contains material resources that are processed by cell and node agents. These resources, which have no autonomy or decision-making ability, are not agents. A resource is a mere passive entity which has only has one piece of information: Its quantity/concentration at each point in the system and in the recent past (see section 3.2.2).

Although agents do not exchange messages about what they do or/and expect, they may still interact indirectly and these interactions are defined as follows:

- Cell agents do not interact together directly. However, a cell agent can release, on its node, resources which may diffuse and be picked up later by other cell agents unknown to it.
- Node agents interact with others when they diffuse resources. To do so, a node agent gathers information about resources levels from its neighboring nodes and decides how these resources are moved around.
- Cell agents and node agents may interact:
    - At a cell agent's initiative, they can exchange resources. This is the case when a cell needs to pick up resources from its environment or release resources into it.
    - At a node agent's initiative, a cell agent status is evaluated and replaced if it is dead. This is a node behavior because it requires information that are not available to cell agents.

Finally, for simulating the impact of the environment of the multicellular tissue (seen also as the "outside world"), an entity is in charge of providing the AMAS with some resources and to remove some others from it. And, obviously, this entity is not an agent since it has only a pure mechanical role and can be seen as the interface between the AMAS and the world around it.

3.2.2 Resources

Resources are at the heart of the system. Cells are "factories" that work on resources to modify them, and use them to survive to continue working on resources to survive, and so on. Although this seems a bit of a pointless circular pattern the interesting part is that it goes against entropy. Even if complex organisms do not represent the majority of species, they exist and this is a feat in itself. Also, some byproducts of this process are quite interesting, like self-consciousness that gives

the Universe the opportunity to think about itself. Anyway, in CoCell, resources are building blocks, energy sources and potential information carriers.

3.2.2.1 Definition

Since we are dealing with the biological world and that at this scale "everything" is chemistry, the first very basic assumption is that resources are molecules. In real life these molecules could be small organic or inorganic molecules, proteins, DNA, RNA, monomers or polymers. In our system, all molecules are considered as resources present in the environment. There is no explicit description of their structure or of their physicochemical properties like composition, reactivity, volume and so on. Neither are there any assumption about their possible/actual role in the system apart for energy sources (see 3.2.2.2).

A set of nR resources R_i which concentrations are $\{[R_i]\}\ with\ i \in \{1, \dots, nR\}$ is defined for the system. We make the distinction of concentrations $\{[IR_i]\}$ that represent the molecules present inside the cells, or internal resources, and the concentrations $\{[ER_i]\}$ that represent the same molecules but present in the environment, or external resources (on the nodes). A real cell does not use all the molecules that are present in its environment. This is why resources present in the simulation only account for the useful or essential subset of molecules of the system being simulated. To add to the genericity of the simulation, all resources are considered identical in their behavior. Only some resources stand out as energy resource that can be used to store energy: The equivalent of ATP for real cells. Although resources are considered equivalent generic molecules, the implementation of the simulation makes it possible to modulate their individual "physicochemical" behavior (like diffusion speed, production cost…) to change the system dynamics or to test some hypothesis.

3.2.2.2 Resources Properties

When a molecule is in a solvent, a very useful property is its concentration *i.e.* the number of molecules per volume unit. Since molecules are small entities and humans usually deal with volumes at their scale (milliliter, liter or cubic meter) the number of molecules per volume unit is huge. In order to manipulate reasonable numbers a new unit was defined: The mole. This unit is expressed by the Avogadro constant, which has a value of approximately $6.022140857 \times 10^{23}$ mol^{-1}. It corresponds to the number of carbon atoms found in 12 grams of carbon-12. Using this unit, concentration is defined as the number of moles per volume unit like mol/l. Concentration is a key parameter in order to study the behavior of molecules in solution.

Concentration of resources in the system are defined at each grid node as well as inside each cell. Since the volume of a node is arbitrary, it is defined as unity and the concentration of a resource is equivalent to its quantity. The ranges of values observed in the simulation for resources are in no way related to actual concentrations of molecules in real cells.

Another property of resources is whether or not they can be used as energy source by cells. In most simulations only one resource is considered as a potential energy source.

Finally, resources can be delivered by the world outside the system or be produced by the system and transported out of it:

a) If a resource comes from the "outside world", it is considered infinite and is replenished every cycle to be available to cells that require it. In biological terms this can be the equivalent to nutriment transport through blood capillaries. Various ways are implemented to provide the system with resources: New resources arrive from one side of the grid; some

discrete sources are randomly scattered or arranged on the grid, or each node receives a small amount of resources every cycle. Only the first case has a strong impact on the dynamics of the system by imposing a gradient of resources from one side of the system to the other. Since these gradients and their impact are not the main focus of our study they are not discussed further and the third implementation is preferred for most of the simulations presented here.

b) If a resource is produced by the system, its concentration everywhere in the system decreases with time. This simulates the fact that it is actively transported to the "outside world". In biology this can be the result of transport to other parts of the organism where this resource is used. The synthesized resources could also be unstable once in the environment and decay with time once they are released from cells. This is particularly true for messenger molecules since they are either degraded or recaptured/recycled by surrounding cells. This drain is a very important property since it limits the range a resource can travel from its release location and the number of cells it can reach. Signaling molecules (apart from hormones) in biological tissues present this feature and it has been shown that the average distance they can travel is around 3-4 cell layers before being destroyed or recaptured (Handly, Pilko, and Wollman 2015; Thurley et al. 2015). This distance value is more or less respected in the simulations depending mainly on the size of the grid. For small grids (below 16×16), a 3-4 cell radius represents a portion of the system too large and local dynamics tend to be averaged out, so the radius must be decreased.

The special non-agent entity dedicated to the simulation of the "outside world" is responsible for the management of resource feeding and removal. Its role is to maintain a constant quantity of matter in the system by providing basic resources (which are supposed to belong to a set named SetA) and removing system specific resources (which belong to a set called SetB). The removal speed is defined for each resource. The total amount of SetB resources removed each cycle imposes the amount of SetA resources that are fed back to the system (Figure 3-3).

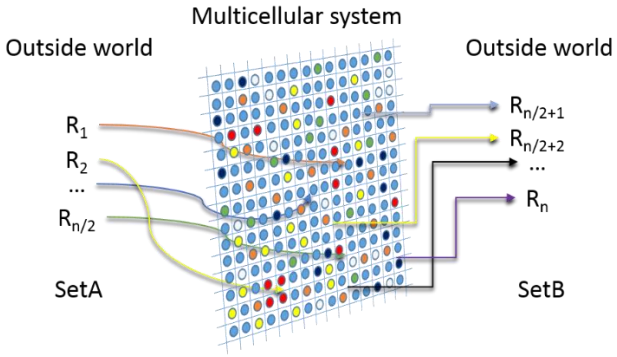


Figure 3-3 Relations of the system with the outside world

3.2.2.3 Resources Dynamics

Two things can happen to resources: Diffusion in the environment (and, as seen before, this is performed by the node agents) and transformation.

Chemistry allows the construction of large molecules from small ones following complex rules. Solvent, reactivity, concentration, geometry and many other parameters dictate which transformations can occur and which cannot. Simulation of these processes is a branch of science by itself requiring costly quantum mechanics calculations. This is far beyond the scope of this work and would probably not add much useful information to our simulation. One strong assumption of our work is that the emergence of communication is not dependent on the nature of the

"substrate" but more on organization. That means that the chemical/physical "details" of the system are supposed to be irrelevant. This assumption is supported by the fact that communication emerged for chemistry-based cells, electro-chemical human brains and even small robots.

So transformation of one molecule into another molecule (resource R_i to R_j) is implemented as a rule-based model. In theory these transformations can occur anywhere (environment or cell cytoplasm) if the conditions are right. But cells are where most of transformations will occur. Thus, in CoCell, resources can only be modified inside cells.

These two mechanisms are described in more details respectively in the node and cell sections.

3.2.2.4 Criticality

The concept of criticality is central to the cooperation process in adaptive multi-agent systems. The notion of criticality is used by a cooperative agent to decide which agent is considered as having the highest degree of dissatisfaction in its neighborhood (Noël and Zambonelli 2015) in order to help it if possible. By designing the right measure of criticality for the agents, the cooperation process can be fully applied. Consequently, the global criticality of the system can be minimized using a local decision mechanism leading the system to perform its intended task.

Real cells communicate by exchanging molecules released in their environment (see section 2.7.4.2). Since we expect this kind of communication to emerge in CoCell, cell agents cannot communicate with one another directly and as a consequence, they are not aware of the degree of dissatisfaction of their neighbors. Actually, cell agents are not even aware that other cells exist at all. All they "know" is themselves and resources.

A cell deals with resources and its survival depends on the availability of these resources. Actually, in the best case scenario, cells only require energy resources or building blocks to produce energy resources. But a cell agent does not know what an energy resource is and cannot bias its behavior towards it. Evolution may have produced safeguard mechanisms that introduce a bias to focus on these particular resources when they are needed; however, in our case we avoided any bias towards energy resources since we try to evolve an early system of cells with basic actions. In fact, it is natural selection that focuses cells on energy production since the ones that do not produce enough of it die soon and cannot propagate their unfit behavior to their offspring.

A cell agent is "satisfied" when it has enough resources to perform the transformations it knows how to do and if it can continue to exist. When it lacks some resources, the cell cannot do the actions that act on them and its behavior is limited. In that case it can be considered "dissatisfied". Thus, the criticality of a cell agent is directly linked to the availability of resources, and becomes a function of the availabilities of all the resources it may need. As a reference to an agent cell criticality and because of its strong relationship with it, the lack of availability of a resource will also be named "criticality" even if a resource is not an agent.

To basically define a resource criticality, let us consider the case where this resource is absent from the system. Since it cannot be used by any cell, this may endanger the whole system, and this resource can be considered as critical. On the contrary, if a resource is always present and available to be transformed by cells, it does not hinder any cell in its behavior and is considered as non-critical.

In CoCell2 and CoCell3, the criticality of a resource will also be modulated by the knowledge acquired by a cell agent about its environment or its internal workings during its life (see next chapters).

There are two sets of resource criticalities: {Critl$_i$} for the criticalities of internal resources {IR$_i$} and {CritE$_i$} for the criticalities of external resources {ER$_i$}. The implementation of the criticality evaluation can vary but in most of the experiments presented here it is composed of the measure of two terms:

- The current available quantity of a resource: Is it rare or plentiful?
- The forecast availability of a resource in the near future: Is it disappearing from the environment or not, and how fast?

A simple implementation example could be:

$Crit_i = \frac{1}{\alpha[R_i] + \beta\frac{d[R_i]}{dt}}$ where [R$_i$] is the concentration of resource *i* (either internal or external) and α, β are constants.

When the resource is nearly absent or disappearing fast its criticality is high. On the contrary, if stocks of a resource are full or being replenished, the criticality of this resource is low.

Either using the aforementioned criticality implementation or another one, a very important property that must be respected is ranking. It must be possible to compare two criticalities and tell which one is higher than the other. From this ranking decisions can be made by the agents.

3.2.3 AMAS Cooperation to Accelerate Parameter Space Exploration

A cooperative algorithm in the AMAS framework can only use local information. The global objective of the system must be carefully left out of any decision process to avoid introducing any kind of bias. This is a prerequisite for our study and the strength of the AMAS approach.

Accelerating convergence is possible through a combination of three elements:

- A random decision strategy at each time step leads statistically to use the possibility of action in an equivalent way: Consequently, a "good" decision can often be overturned by a "bad" decision that follows it. The cooperative decision has regularities because the same environmental context will lead to the same decision; this does not yet explain the accelerated convergence.
- Cooperation is a proscriptive method that cuts out possibilities and almost systematically leads to a single possible decision for a given situation. Thus, there is no exploration of research space as usual in heuristic or complete approaches. Acceleration is the immediate consequence of traveling through this tiny subspace; even if it does not explain why the "right" solutions are there yet.
- This tiny subspace contains, by construction, global states where the components that make up the system have permanent cooperative interactions. This functional adequacy property has been proven in (Camps, Gleizes, and Glize 1998) which states that any system in permanent cooperative interaction is functionally adequate in its environment. Here the system is made up of all the cells, while the environment is the tissue that contains the materials and laws of the environment.

3.2.4 Node Agents

3.2.4.1 Definition

In real multicellular tissues, resources diffuse in a liquid medium that is reproduced in CoCell, as a grid of node agents. This grid is the environment where cell agents live and is distinguished from the environment of the system named "outside world" that provides basic resources and harvests complex resources. Each node in this grid has a predefined number of neighbors which can be defined arbitrarily. However, in a 2D-grid, a node has generally 4 or 8 neighbors, as depicted in Figure 3-4.

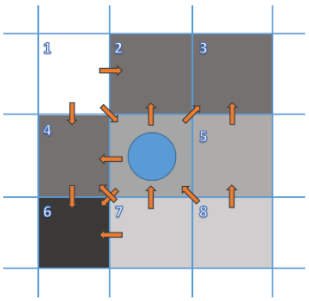


Figure 3-4 Neighborhood of a node and diffusion of resources (arrows from white to black). A cell living on the node is represented in blue.

The role of the environment nodes is threefold: Perform the circulation of resources around the cells, evaluate the survival status of cells and replace dead cells with mutated daughters of a selected neighbor cell. The first two behaviors are performed every cycle and do not interfere. The third one is dependent on the result of the cell survival evaluation behavior and is only triggered if the cell is found dead. Thus, these three behaviors are described in the next sections with their corresponding Perceive-Decide-Act life cycle.

3.2.4.2 Resource Diffusion

Diffusion in the medium is the balancing of concentration over time and space. Typically, if a resource is more concentrated in point A than in a neighboring point B it will tend to diffuse towards B until concentrations in A and B are equal. Under the assumption of steady state, an exact solution of this process is given by solving Fick's laws of diffusion (Crank 1975). The implementation of these equations is part of the Ordinary Differential Equations (ODE) and Partial Differential Equations (PDE) Models algorithms described in sections 2.9.3.5 and 2.9.3.6. However, this system of equations is computationally costly to solve and an accurate depiction of the diffusion phenomenon is not crucial to the study given the simplifications in other parts of the system. Another way to perform diffusion without solving differential equations is to use a multi agent system as described in (Redou et al. 2007). Here we use the same principle of a multi agent system although the inner workings of the agents are different. So, a very simple algorithm is used to implement the diffusion.

Following the classical agent life cycle terminology, the behavior of a node agent is performed as follows:

- Perception: At time t, for each resource ER_i the node perceives the quantity present in its neighboring nodes.
- Decision: The node compares the quantity of each resource with the quantity present in its neighboring nodes. If the quantity of the resource i is inferior to its quantity on a neighbor n,

nothing happens with this neighbor. If the quantity is superior, the node proposes to transfer half the difference to n. In theory, this would be enough to balance the resource i between the two nodes. In practice, other nodes might also need resource transfer and the quantity to transfer might be too large to be diffused in a single time step. So, the node evaluates the maximum quantity it can diffuse to each neighbor to keep ERi superior or equal to those on these neighbors. Finally, each quantity is capped by the maximum quantity of resources diffusible in one cycle.

- Action: The calculated quantity of resource is transferred to each neighboring node.

3.2.4.3 Cell Status Evaluation: Viability

Every simulation cycle, each node checks if the cell present at the same position is still alive. This evaluation is not performed by the cell agent itself since it requires information that are not considered accessible by the cell. There are several ways for a cell to die, so various pieces of information are evaluated to decide whether or not the cell can live another cycle and perform its tasks.

Here are some possible reasons for the death of a cell in the simulation:

- Old age: Some mutations accumulated during the life of a real cell may impact its viability. When too many bad mutations alter the behavior of this cell, it can decide to self-destruct by initiating the apoptosis process. To take this phenomenon into account in the simulation, the probability of spontaneous death is proportional to the age (expressed in simulation cycles) of a cell agent.
- Lack of energy resource: A real cell can use one or more resources as energy source for its normal functioning. If the potential energy reserve is evaluated to zero, the cell is considered dead. However, to prevent a cell agent from taking biased decisions towards the optimization of special energy resources, it is unable to know which resources are used as energy resources. Therefore, a cell agent is unable to know when it dies by lack of energy.
- Toxic compounds or pathogen invasion: A toxic compound is deleterious because it blocks or perturbs the regulation of some crucial processes inside real cells. For example, it could specifically bind and inhibit a protein that is essential for ATP production. To resist the influence of a toxic compound, a cell either needs a bypass mechanism that can perform the same function as the blocked pathway or destroy the toxic compound. In either case, something needs to be manufactured by the cell to react to the threat: Either an enzyme to metabolize the toxic molecule or a set of protein to form the bypass. A pathogen can block pathways, reroute cell processes to its own uses or deplete important resources. As for the toxic compound, the cell needs to react by activating relevant defense mechanisms. This usually translates into molecules that will kill the pathogen or trigger the immune system. For the simulation, these two threats are processed in the same way, their presence must be answered by the production of one or more specific user-defined resources and their concentration maintained above a given threshold. If these internal resources are below this threshold, the cell is considered dead. Since we do not want to give the cell the notion of what is a threat to its survival, this evaluation cannot be performed by the cell agent.

So, before a cell agent is allowed to perform its life cycle, the node, on which it resides, determines if it is still alive in the following way:

- Perception: The node queries the cell for the level of resources it possesses and in particular the resources that can be used as energy source. The age of the cell is also perceived and whether or not the cell has performed an apoptosis action.
- Decision:
    - As a cell gets older it has an increasing probability to die of "natural" causes.
    - If the quantities of resources that can be transformed into energy are zero, the cell is dead.
    - If a pathogen or a toxic compound is present in the cell, it might be killed. If the set of resources required to resist are below a given threshold, then the cell is dead.
    - In some cases, a cell can decide that it is best for the community if it dies. It that case it has performed an apoptosis action the previous time cycle and it is dead.
- Action: The cell content is released on the node or destroyed depending on the death type. It is then flagged to be processed by the dead cell replacement task.

Once this evaluation is complete, a cell is considered alive or dead (there is no in between state). If it is dead, the node will activate its dead cell replacement behavior but an issue remains: Whether to release the dead cell content in the environment or not. In real life, both scenarios can take place depending on the circumstances. Programmed cell death is a complex process that includes a kind of tidying up step where potential harmful compounds are inactivated before release in the environment. On the contrary, brutal cell death due to unforeseen circumstances usually release the full cell content into the environment. This can be sometime beneficial since compounds that should stay inside the cell are present in the medium and can signal that something is wrong somewhere to other cells and to the organism defense system.

In the simulation context, the release of the content of a dead cell is subject to discussion. Indeed, only important resources are simulated and they do not include toxic compound for other cells. Hence, upon release they can be immediately useful to neighboring cells. In this fashion, cell death can be seen as a massive "production" mechanism. This can promote the apparition of "cheater" cells that feed on resources released by dead cells and do not benefit the system since they do not require to participate in the network of interdependencies. Although these cells can exist in real life, we are not interested in their presence in the simulation since they would not participate in the communication. So, apart in very specific situations and experiments, all cell death is clean and rare resources from SetB are transformed into SetA resources before release or simply discarded from the system.

3.2.4.4 Dead Cell Replacement

If a cell is found to die, its corresponding node has the very important task of replacing it. This process is described in details in section 4.2 but the following points are important to mention at this stage:

- The new cell must be the result of a mother cell dividing into two daughter cells. One daughter will stay at the position of the mother and the second one will be placed at the dead cell position.
- The mother cell must be located at a reasonable distance from the daughter cells. The migration process from the mother location to the final position is not simulated since, as mentioned before, motility and chemotaxis (movement in response to a chemical stimulus) are not considered relevant to the study.

- The resources concentrations inside the daughter cells are the same as in the mother cell. This is subject to debate and is discussed in the next section.
- Daughter cells have a chance to be mutated. Mutations can actually occur any time during the lifespan of a cell. But in terms of implementation and analysis it seems more suitable to perform this task on new cells. The impact of this choice on the simulation is not considered important, given the average age of cells that is small compared to the system age.
- Selection of the mother cell must be more efficient than random.

The last criterion is of particular interest. Indeed, Nature uses randomness and natural selection to evolve a system. Its strength lies in the number of parallel systems being evolved and the available time. Billions of billions of cells changing, surviving if they are fit to their environment during billions of years can lead to multicellular organisms that we see today. But in our simulation we have nothing near this processing power. So we must somehow orient the process towards a higher degree of organization without introducing to much bias on the phenomenon we want to observe *i.e.* cell communication.

This task of dead cell replacement, is performed by the node agent during its classical life cycle:

- Perception: The agent queries the neighboring nodes for their resource content and variations on a time period. Then the node queries potential mother cells in a certain radius around it for the resources quantities they gathered and released on a time period.
- Decision: From the capacities of each potential mother cell to fit the requirements of the node neighborhood in terms of resource, a candidate is selected. This is detailed in section 4.2.
- Action: The mother cell is duplicated at the node position. Then, following a probability function, the daughter cell at the node position is requested or not to mutate.

3.2.4.5 Cell Resources During Division

In real organisms, the quantity of each resource held by a cell is equally split between its daughters when this cell divides, but at the same time their volume is half their mother's. Consequently, the concentration of resources is the same as before the division process.

In CoCell, at the end of the dead cell replacement process, a copy of the selected cell agent (the mother) is placed on the dead cell node. Since the cell volume is not defined, resource concentration and quantity are equivalent which leads to a problem for the division process. Indeed, duplicating the resources correspond to an artificial increase of the total resource quantity in the system and it is possible to imagine situations where this mechanism alone is sufficient to allow cells to survive.

On the other hand, dividing the resource quantity equally between the daughter cells, while maintaining the global quantity of resources introduces an even stronger side effect. From a nicely working mother cell with internal concentration $\{[IR_i]\}$, we switch to two daughter cells containing $\{[IRI_i]/2\}$. In terms of parameter space of the cell, this is equivalent to a drastic translation into a poorer region. The internal mechanisms of the cell are unlikely to be able to deal with such changes and even if they can, the repercussions can have a major impact on the system as a whole. For example, the cell can completely switch its behavior under these new circumstances with the direct consequence that the daughter which replaced its mother at its location will not provide the neighborhood with the same services. Consequently, the dynamics of the other cells will change. In some cases, the local group of cells might be resilient enough to readjust to the new situation but it might also lead to cell death and system failure.

Furthermore, when a cell agent is selected to replace the dead one, this can only be based on its behavior in the past. This behavior is completely dependent on its resources content. A cell with half the resources would not have behaved in the same way and might not have been selected as a good replacement.

This problem could be avoided by introducing a notion of cellular volume in the simulation and a specific model for cell growth. This would add different kinds of problems to the simulation. It is not the solution selected in CoCell although it would be interesting to investigate in future works.

Here the choice is that daughter cells have the same resource content as their mother. The system is monitored for situations where this mechanism increases the global resource content and the entity that regulates the exchanges with the outside world compensates for it. The fact that cells do not release their inner content in the environment upon death is another compensatory mechanism.

3.2.5 Cell Agents

Cells agents (Figure 3-5) represent the focus point in the system since we are mainly interested in the dynamics of their communication protocols. They are not more important than the rest of the system, since they cannot exist without it, but they represent the only elements that can modify their behavior and adapt to changing conditions.

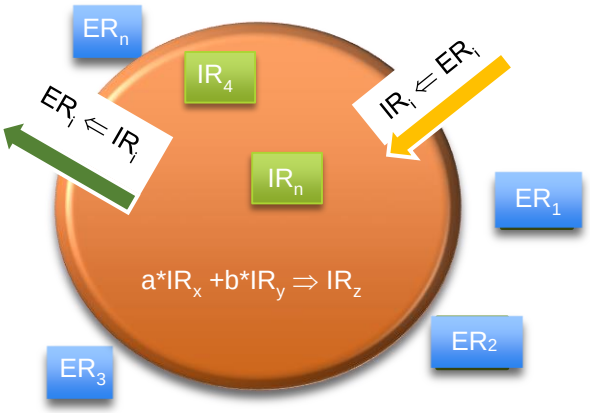


Figure 3-5 Schematic of a cell agent

As mentioned in section 2.9.3.1 several attempts have been made and are ongoing to simulate "complete" cells either alone or in communities. These efforts are extremely time consuming and require the best possible supercomputers available. The main problem is the number of parameters. There are roughly 20,000 genes in a human cell. Even if all of them are not used and transcribed simultaneously, this means that at any given time a cell will have thousands of active proteins, with possibly as many small molecules species. All may have different concentrations and different localizations. Since we have not characterized in details all the possible proteins present in a cell and in particular their catalytic potential and constants, it becomes challenging to perform an accurate simulation of this complex system in action even using accurate laws of physic. Furthermore, to the best of our knowledge, the exact content of a single real cell has never been established for obvious methodological difficulties. Even if it is now possible to follow a single protein in a cell or to know protein expression patterns, metabolites concentrations are only known as averages on large populations and their list is probably not exhaustive. These lacks of data inevitably lead to questionable simulations since the cell is clearly a complex system and wrong initial conditions are as damageable to the results as a poor model.

Given the current limitations of biological knowledge and available data, we did not consider trying to implement even a simple model of cell. Indeed, even with all available data such an accurate model would require many PhD theses to complete.

Since our interest lies in the communication between cells and not really in its inner workings (gene regulation, transport, compartments, proteins, metabolites…) we choose to use the simplest cell model possible that is necessary for the emergence of communication.

Even if this cell model is crude it is sufficient to form a complex system with emergent properties as described in the next chapters. If evolved communication protocols are part of the emergent properties, then the simplicity of this model becomes an advantage since the findings are valid for a much larger class of systems than living cells.

3.2.5.1 Definition

In CoCell, a cell is an agent since it is an autonomous entity that has the ability to take decisions and act on its environment based on local perceptions. To do so, it has access to some aspects of its internal state and to data about its immediate environment. Its actions only impact itself and its close neighborhood. It can die and in some circumstances divide. The implementation of the cell function is based on rule-based modeling.

A note of importance about division: We consider that the simulated "tissue" is in a steady state where cell division is at a minimum and only occurs as a repair mechanism to replace dead cells. This is in contrast with many multicellular simulations where division is a process that occurs whenever some conditions of resources and/or size are met. Usually these simulations deal with pathological cells like those found in cancer which is not our focus here.

3.2.5.2 Cell Properties

A cell agent is defined by:

- Its position in the system. This is important to define a set of neighbors and for resource concentrations in the corresponding environment node. Cell mobility is not simulated so a cell exists its whole life at the same position.
- Its resources content. There is an upper limit to what the cell can contain for each resource. This internal resource content influences what the cell can do in terms of transformations at any given time.
- Its energy resources. These resources are consumed whenever the cell performs an action. Each of these resources can have different energy yields. In most of the simulations presented, all cells have the same single resource to produce energy. An important thing to remind is that a cell is not "aware" that a resource can provide energy. This information must not be used during the decision process to avoid bias towards energy optimization.
- A set of actions {Ai}. This set dictates what actions a cell can theoretically perform. Resource availability as well as the internal decision process activates proper actions. They can be of three types:
    - Gather action (gA). A gather action transfers resources from its corresponding environment node to the cell interior (cytoplasm). Each gather action can only transfer one type of resource. The cell will gather from its environment node and neighboring nodes. This can be formalized as:

$gA_i: ER_i \xrightarrow{gA_i^{Ecost}} IR_i$ where gA^{Ecost} is the energy cost for this action and i the index of the transferred resource

- Release action (rA). A release action transfers resources from the cell cytoplasm to the environment. The release occurs on the cell node and all its neighboring nodes, with a cost rA_i^{Ecost}

  $rA_i: IR_i \xrightarrow{rA_i^{Ecost}} ER_i$ where rA^{Ecost} is the energy cost for this action and i the index of the transferred resource

- Production action (pA). A production action transforms a set of internal resources into another single resource. It can be written as:

  $pA_j: \sum_i a_i . IR_i \xrightarrow{pA_j^{Ecost}} IR_j$ where pA^{Ecost} is the energy cost for this action, a_i are stoichiometry parameters for the reaction and $\sum_i a_i = 1$

  A production action can correspond to a single reaction or a complete production pathway. In the latter case all intermediate resources are considered as background materials and are not explicitly listed in the resource set.

  Interrelations between productions as observed in real cells, like feedback loop mechanisms and agonism or antagonism, are voluntarily left aside. This is because in a real cell, these interactions represent the actual mechanism that will drive the decision process. In the simulation, this would overlap and possibly antagonize the cooperative decision process that we intend to implement based on the AMAS paradigm.

- Apoptosis action. The cell node agent commits suicide and needs to be replaced. This can happen for different reasons like uselessness. This action is signaled to the corresponding node age and triggers the "dead cell replacement" behavior.

The amount of resources altered per cycle, R^{Δ}, is the maximum between an upper limit max^{Δ} and the minimum available antecedent resources (reactants) of the action: $R^{\Delta} = \max\left(max^{\Delta}, \min([R_k])\right)$, in other words, the cell will always try to work on the maximum available resources. Each action consumes $A_i^{Ecost} * R^{\Delta}$ energy units when it is executed.

This is the general layout of a cell agent used in the CoCell system. Nevertheless, in some instances of the simulation, it is modified or adjusted to test some hypotheses; and the modifications made to this model will be mentioned when necessary.

3.2.5.3 Life Cycle

A cell agent performs a classical Perceive-Decide-Act life cycle. At the beginning of each cycle, if its node agent evaluates it is still alive, then the cell agent is allowed to perform its three phases as follows:

- Perception: The cell updates the quantity of resources IR_j that are available internally to determine what it may use. It also updates its information about the external resources ER_j that are available on its corresponding node.
- Decision: Based on its perceptions, the cell evaluates if the conditions to execute a given action are met for listing all the actions it could possibly perform during this cycle. For example, a release action $IR_i \rightarrow ER_i$ can only be executed if IR_i is present inside the cell and/or above a given threshold. From this list of possible actions, the cell then selects the best appropriate action based on its local goal and cooperation towards others and itself.

This decision process is an abstraction of all the chemical mechanisms and feedback loops present in the cell to handle its next action.

- Action: The cell then executes the selected action and updates its resources content accordingly.

A precise description of this life cycle is done in the next chapters for each instance of CoCell.

3.2.5.4 Cell Mutations

Mutation is the process that introduces novelty into the system and allows adaptation. Real life mutations are random and can sometime be propagated to offspring but most of the time they only affect a cell and alter its behavior (positively or negatively). As mentioned in section 2.7.2, the random process is good enough when near infinite time and test subjects are available. But in the simulation, computing resources need to be optimized. In this situation, fully random changes are not possible. Then the question about mutations becomes: How to change an individual in order to improve its fitness to its environment without introducing too much bias towards the goal of the simulation? The answer given by the AMAS approach is cooperation based on local perceptions. Thus, the implementation of this behavior of the cell uses cooperation to decide the best mutations to perform based on local information. This form of mutation is to be compared with prokaryotes behavior where DNA fragments can be exchanged between cells to acquire new functionalities (Griffiths et al. 2000). However, to avoid any bias, each cell has to autonomously decide which mutations to perform, in order to change its behavior, without any knowledge of the goal of the global system. Mutation algorithms are presented in the chapter 5 CHAPTER 5describing CoCell2.

3.3 Agent Scheduling

All agents, nodes and cells, can potentially act on the same resources at the same time. Thus, it is important to design a scheduling pattern that can solve possible conflicts. The easiest way is to use two independent grids for the external resources: the "current" grid and the "future" grid. Agents take their decision based on the resource concentrations in the current grid and their actions take effect on the future grid. In this way, the result of the actions of an agent A on the decision of an agent B are the same whether it is acts before or after B.

There is still a problem when two agents, A and B, try to gather the same resource R at the same time from the same node. If the sum of the amounts A and B try to gather is higher than the actual content of the node, a problem arises. To solve this, a two-step mechanism is used: In the first step, A and B queue a request for R and the amount required. After all agents have acted, requests are evaluated and if the total amount is over the content of the node, each agent receives an amount proportional to its requested amount divided by the total amount requested by all the agents. This solves any possible conflict of agents accessing the same resource.

3.4 CoCell Simulation Platform

All the implementations of the AMAS simulations are done in the CoCell platform. This platform is a program developed in C++ that uses the Qt (https://www.qt.io) library for the human interface and OpenGL (https://www.opengl.org) for custom graphics. The dependencies are cross-platform making CoCell compatible with all popular platforms. The main target is the Windows system but it was also compiled on Linux to be executed offline on a cluster. The application takes advantage of multi-core processors when possible. In particular, the decision phase which is performed individually for each cell and is a local operation is parallelized. Thus, a simulation of 100×100 with

40 resources can be managed on an average desktop machine at 1-5 cycles per second (slightly faster in offline mode).

Typically, an interactive CoCell run consists of two windows. The main window (Figure 3-6) displays the system space with cells as squares or cubes depending on the 2D or 3D setup. Cell colors can represent any of the features selectable on the side of the graphic frame. The color range is updated dynamically and the color Look Up Table (LUT) is chosen to maximize the visual differentiation of zones. Information about either cells or nodes can be displayed.

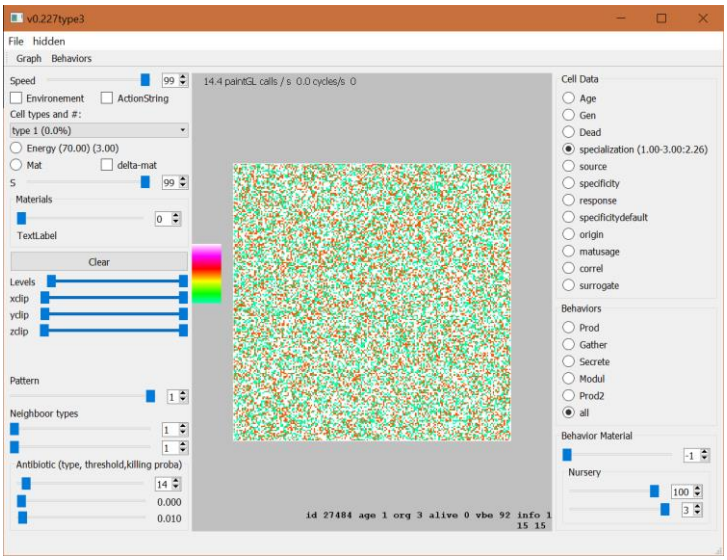


Figure 3-6 Main CoCell platform window. Cells on the grid are represented as squares colored according to one property.

The second window (Figure 3-7) displays the same information as the main window but in form of histograms to show the distribution of values in the system or the average value (with min and max) as a function of time to observe the dynamics and history of the system.

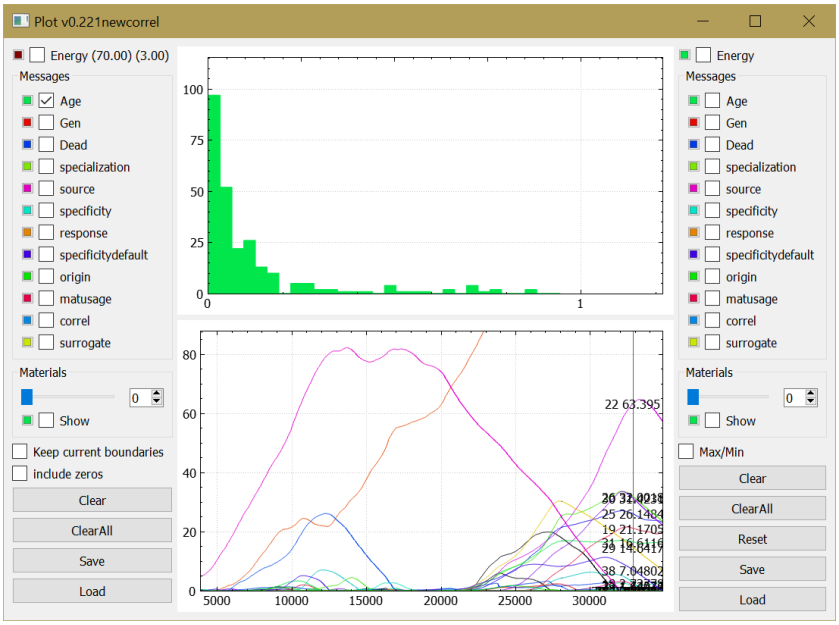


Figure 3-7 Statistic window to survey temporal evolution of cell properties or distribution of values.

Three information windows can be summoned to display the state of any particular cell in the system under the mouse cursor. The first one (Figure 3-8) displays the resource correlations

observed by the cell as well as the activated actions and criticalities of each resource. It can also display the actual cell decision process in details at each time cycle.

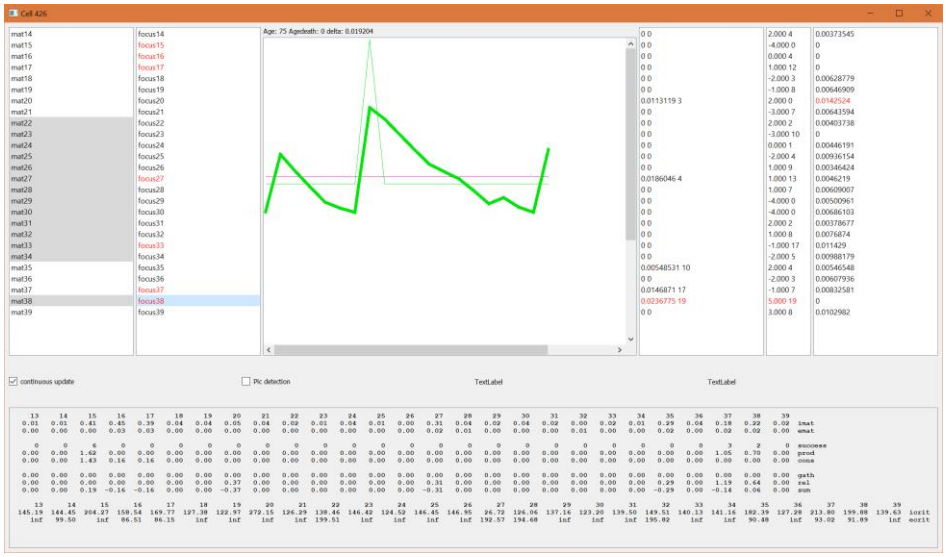


Figure 3-8 Cell information window. Correlation can be displayed real time as well as the decision process at every cycle.

The second information window (Figure 3-9) shows the actions that are available to the cell, with the reactants, products, factors, energy cost and the number of times it is used. Resources usage is also displayed.

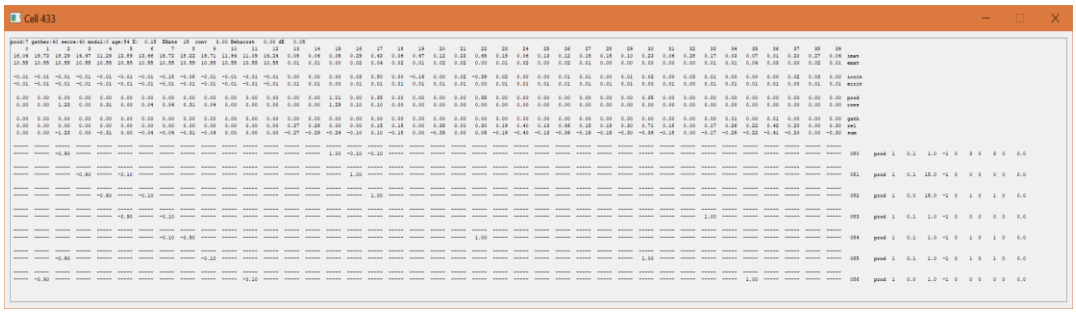


Figure 3-9 Cell actions window. Some or all actions are displayed with their components, cost and usage. Resources information are also shown.

The last information window (Figure 3-10) displays the neighborhood of the pointed cell. Various information about the neighbors can help understand the dynamics of the cell and its relations with them in terms of dependency and cooperation. This window is only usable in the case where there are 8 or less neighbors per cell.

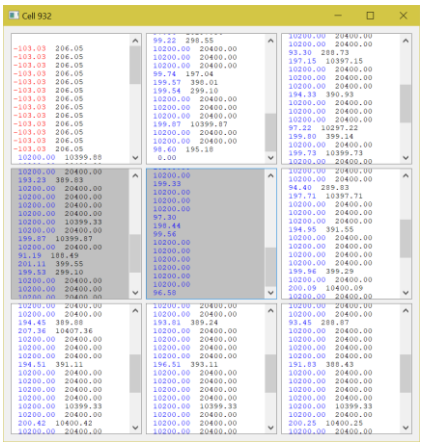


Figure 3-10 Neighborhood information of the currently selected cell. Various information can be displayed.

The behavior of any cell can be observed in real time or recorded for further offline analysis. It is also possible to record multiple cells to analyze their interactions in more details.

The offline analysis is done in an auxiliary application: TraceView, written in C++ and using the Qt library. The main window (Figure 3-11) displays variation of selected variables with time, the second is used to select the information to display and the third shows the decision process of the cell for the cycle pointed in the main window. This analysis is very useful to identify bugs in the cell agent design and understand why a cell dies or why it executes one action or another.

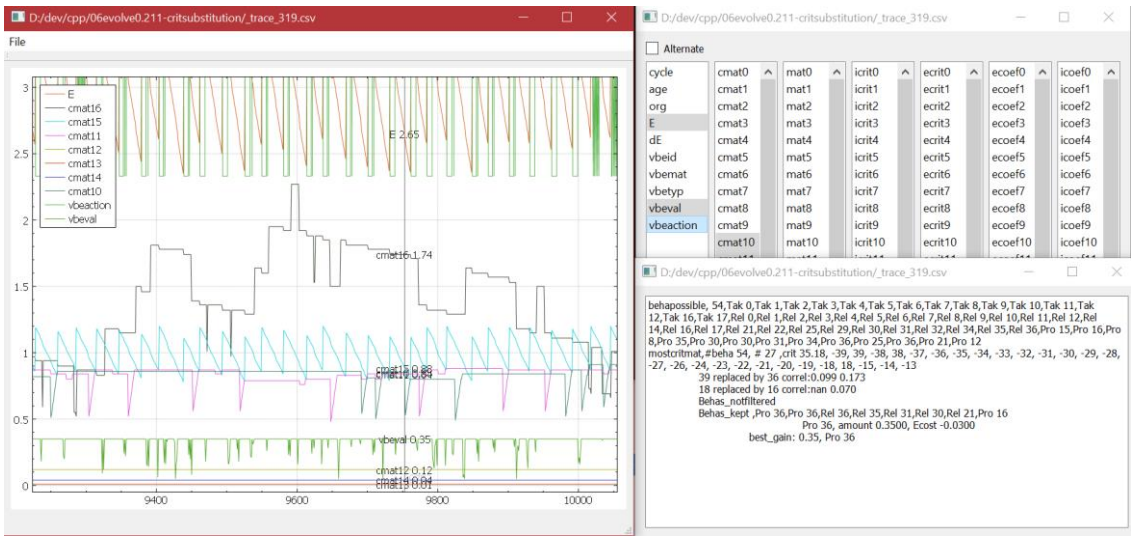


Figure 3-11 TraceView windows help analyze the cell behavior in details and offline.

3.5 Conclusion on the CoCell Model

In this chapter we have defined the different parts of our system, their roles and ways to fulfil them. This includes two types of agents: The nodes and the cells. These agents exchange resources that are used by cells to survive and hopefully communicate with one another. The node agents have the role of circulating resources and replacing failed cell agents using local information. The cell agents take cooperative local decisions to either collect, release or transform resources. Finally, an independent entity is used to feed some resources in the system and remove others to maintain the consistency of the biological environment.

From the informatics standpoint this thesis tries to understand the conditions for self-organization of multi-agent systems when communication is not formalized nor explicit and information about neighbors is limited. Each cell is only aware of its immediate environment but to take decisions it uses two powerful assumptions that help the cohesion of the system. Namely, the other cells are also cooperative and they work and take decisions in the same way. This is related to the notion of kinship and is a very strong driving force in the development of multicellular organisms. Although communication through the environment has been studied like for ants using pheromones or robots using lights, the difference here is that cells do not know that a given resource is a signal or just a building block. The resource role assignation must emerge from the cooperative behavior of the cell community. In this regard, AMAS is used in a context that is not usual since in most situations a scientist will add as many local information as possible to make the simulation more likely to converge. Studying the conditions of self-organization in this difficult setup is an interesting case for the extension of the AMAS application domain.

The second aspect of interest is to show that cooperation between entities speeds up the convergence toward solutions from an initial random process. This will be proved by the capability to find solutions in reasonable time in gigantic search spaces using standard personal computers.

The third point concerns the potential bias that cooperation could induce in finding solutions. It is currently used in the AMAS as a solving principle that differs from standard approaches because this decentralized bottom-up principle leads to emerging global solutions in the sense that no agent is guided by an evaluation function of these solutions. But never before was considered the implicit and invisible bias that cooperation would carry towards converging solutions in a particular subspace of the problem. Our results would be distorted if this were the case, but we will show experimentally that such a bias does not exist.

In the next chapters, we instantiate this system and perform experiments to test various hypotheses. Three stages are described: In the first one, we test the cooperative cell agent decision-making and the cooperative node agent cell replacement. In the second stage, we introduce a cooperative mutation process. Finally, in the third stage we study the conditions for communication to emerge between the cell agents.

CHAPTER 4. COCELL1: REACHING A DYNAMIC EQUILIBRIUM

Ce chapitre présente la première étape vers l'émergence de la communication entre cellules simulées. Sur la base des choix et des concepts exprimés dans le chapitre précédent, l'objectif du travail présenté ici est de faire en sorte que le système multicellulaire atteigne un équilibre dynamique lui permettant de survivre. Cette survie est un moyen d'évaluer la pertinence des choix faits précédemment.

La première partie de ce chapitre instancie et complète les comportements des agents cellulaires et nodaux présentés précédemment. Notre but ici est de doter les agents d'une attitude coopérative, ce qui signifie aider l'agent le plus critique de son voisinage (y compris lui-même) sans rendre les autres plus critiques. Selon la théorie AMAS, la coopération est le moteur de l'auto-organisation et de l'émergence d'une fonction collective globale et adéquate (ici la survie de types de cellules complémentaires). Ici, la coopération est ajoutée pour qu'un agent cellule décide quelle action doit être exécutée et qu'un agent nœud décide quelle cellule remplacera une cellule morte.

Ensuite, ces comportements coopératifs sont testés sur différents scénarios afin d'étudier leurs propriétés telles que la stabilité, l'efficacité et l'évolutivité, et de discuter des possibles biais.

This chapter presents the first step towards the emergence of communication between simulated cells. Based on the choices and concepts expressed in the previous chapter, the objective of the work presented here is to make the multicellular tissue reach a dynamic equilibrium enabling it to survive as a whole. This survival would evaluate the relevance of the choices previously made.

The first part of this chapter then instantiates and completes the behaviors of the cell and node agents presented previously. Our point here is to endow agents with a cooperative attitude which means helping the most critical agent in its neighborhood (including itself) without making others more critical. According to the AMAS theory, cooperation is the engine of self-organization and emergence of a global collective and adequate function (long survival of complementary cell types). Here, cooperation is added to make a cell agent decide which action has to be performed and a node agent decide which cell will replace one which died.

Then, these cooperative behaviors are tested on different scenarios in order to study their properties such as stability, efficiency and scalability, and discuss possible biases.

4.1 Cell Agent: Cooperative Action Selection

As described in the previous chapter (see section 3.2.5.3), while a cell agent stays alive, it performs a Perception-Decision-Action life cycle. The description made below completes the one presented in the previous chapter by adding the cooperative attitude required to make a cell agent, a cooperative one.

4.1.1 Perception

As seen before, in this phase, the cell integrates all information available about its internal state and its immediate environment (its corresponding node).

However, the satisfaction of a cell agent is directly linked to the availability of the resources it may need. Therefore, in practical terms, a cell perceives the criticality of each resource IR_i in its internal medium (internal resources) and in the environment of its node ER_i (external resources). Once sorted, these criticalities give the cell a measure of which resources need to be attended to and where.

4.1.2 Decision

As mentioned in section 3.2.5.2, a cell typically has a set of gather actions $\{gA_i\}$ (one for each resource), a set of release actions $\{rA_i\}$ (again one for each resource) and a set of production actions $\{pA_i\}$. To be able to activate a given action some conditions have to be met. For example, a production action can only be performed if all the necessary reactants are available inside the cell. Actions are designed to not only depend on the resources they acted upon but also on other resources that could act as modulators. These modulators can deactivate an action if above or below a threshold, or if outside a certain range of concentration. This is done to simulate the interdependency of pathways and associated feedback loops as observed in real life cellular pathways. But in most of the experiments presented, this mechanism is disabled to avoid interference with the cell agent decision-making process.

So, first, a cell agent lists all the actions that can be possibly executed considering its perceptions. Among all these possible actions, there are many ways for the cell to decide which action to enact based on its perceptions.

The most basic way is random. Obviously this mechanism is not very efficient and the system in this case has little chance of reaching any kind of dynamic equilibrium in a reasonable time. More elaborate methods could be based on the actual system used by real cells. Doing so is difficult for many reasons. One of them is that we do not have yet all the static and dynamical data about living cells in action. Another one is the computational cost of such an implementation even if abstracted to a set of rules. Finally, nowadays cells possess well established communication protocols. It would therefore be difficult to leave out this part of their normal function to observe its emergence without affecting all the other modeled processes.

For the reasons mentioned before, a cooperative heuristic is used for this decision-making. Since cooperation is also an efficient tool to explore the parameter space of a system, it is expected to speed up the convergence toward an emergent solution.

Once the set of executable actions $\{eA_i\}$ is established, the cooperative decision process is as follows (see Algorithm 1 below):

- Both external and internal resource criticalities are put in a single list and sorted.
- Evaluate the impact of all $\{eA_i\}$ on internal and external criticalities.
    - For example, a release action rA_j transfers an internal resource IR_j to an external resource ER_j. Thus, IR_j decreases and its criticality increases. Conversely, the quantity of ER_j increases and its criticality decreases. If this action rA_j is cooperative, the sum of all resources criticalities involved should be negative and thus decrease the cell agent criticality.
- From the most critical resource to the least critical one.

    - Select actions that can improve this resource criticality but without degrading more other critical resources (if any).
    - If no action can impact favorably this resource, select the next most critical resource.
- If more than one action remains at the end of the loop, select one randomly using the criticality gain as its probability.
- A single action is now ready to be executed.
- If no action can be selected for several consecutive cycles, the cell considers itself useless in the system and performs an apoptosis action that will signal its corresponding node that it died.

This base algorithm for the cell decision is adapted in some of the experiments, and the modifications made are explained in the corresponding section.

Algorithm 1 Cooperative Action Selection Process
$\{eA_i\} \leftarrow$ Set of all actions that could be performed
$\{crit_j\} \leftarrow$ Compute resources criticalities and sort them in this set
Compute impact of each action in $\{eA_i\}$ on resource criticalities
WHILE {an action is not set} do
    Consider the next most critical resource in $\{crit_j\}$: R_j
        . Sort $\{eA_i\}$ according to the impact on R_j
        . Select the action(s) (if any) with the most positive impact on R_j and least negative impact on more critical resources (if any)
ENDWHILE
IF {more than one action is selected}
    Choose an action randomly using resource impact as probability
ENDIF

4.1.3 Action

The action selected during the decision phase is now ready to be executed by the cell.

A release action transfers some internal resource to the cell node and to all neighboring nodes.

A gather action tries to collect a given resource from the cell node agent and neighboring nodes agents. Since this process is concurrent with neighboring cells gather actions, it is not possible to know at the decision time if the amount of resource actually gathered is the one required by the decision process (based on impact on criticality). Thus, in some situations it is possible that the selected gather action is not the optimal solution. Again, gathering from neighboring nodes agents coupled with release on neighboring nodes is in some way an active diffusion mechanism that could interfere with the passive diffusion in some circumstances. So this can be replaced by gather and release actions on just the cell corresponding node agent.

4.2 Node Agent: Dead Cell Replacement (Nursery)

When a cell dies, it becomes useless. In the sense of the AMAS theory, uselessness is a Non Cooperative Situation which disrupts the local cooperative behaviors of cells. In order to restore cooperation, a new cell has to be selected to replace the dead one and it has to be more adapted to the local environment than the previous one in order to prosper and the whole system with it.

This replacement is done by the node agent on which the cell died and was expressed in the previous chapter (see section 3.2.4.4) in terms of its Perception-Decision-Action life cycle. During its decision phase, a node agent has to decide which mother cell is better suited to give birth to a daughter which will replace the dead cell. This process of selecting a better suited cell to be born on the now empty node is called the "nursery". This nursery is very important in order to accelerate the exploration of the systems' parameter space. By selecting the "best" cell to replace a dead one instead of a random one, the dynamic equilibrium must be reached faster and be stable. A very strong condition for the nursery process is that it must not include bias towards the goal of the system that is communication between cells. Also, the process must be based on local information since in real life a cell will decide to initiate division based on local environment cues and its internal state.

Several strategies can be applied by a node agent to decide which cell candidate is the most adequate to replace a dead one. For example, a nursery may consider different criteria of choice based on the age of the cell candidates, their energy level, both age and energy, the resources they produce, the similarity of their surroundings, and so on.

Therefore, one of the aims of this first version of CoCell is to devise and compare three types of nurseries among the most representative ones: 1) Nature's favorite: Random nursery, 2) altruistic and 3) cooperative nursery. These nurseries are presented hereafter before giving results of experiments using them in the next section.

4.2.1 Random Nursery

This is the simplest way of deciding: From the selected candidates pick a random one. It works quite well in a large environment like Earth but fails miserably in a simulation with limited resources. Its main interest is that if the random generator is good, this method is assured to be completely unbiased. In theory it could be used in order to compare the effectiveness of other algorithms in accelerating the convergence of the system and to evaluate a possible bias.

Intuitively, the random approach is unlikely to provide a good way to evolve such a simple system as ours since any advantageous cell replacement that is performed has an equal chance of being removed in the following generations. With this kind of nursery, there must be some kind of mechanism to keep good replacement in the long term.

4.2.2 Altruistic Nursery

This nursery tries to balance the resources requirements at the location where the new cell will be born with the resource productions/releases of the candidate cells. In order to do so the decision process is as follows:

- For each resource R_i, calculate the mean value of external variations over a time window Δt and over all dead cell neighboring nodes.
  - $\overline{d[ER_i]^{neighbors}} = \frac{1}{\#neighbors * \Delta t} \sum_{neighbors} \sum_{\Delta t} \Delta[ER_i]^t$
- For each resource R_i, calculate the mean value of resource variations over a period of time for the candidate cell j.
  - $\overline{d[ER_{ij}]} = \frac{1}{\Delta t} \sum_{\Delta t} \Delta[ER_{ij}]^t$
- For each candidate j, evaluate the match between neighborhood resource requirements and candidate resource releases.

- $Score_j = \sum_i^{resources} \left| \overline{d[ER_{ij}]} + \overline{d[ER_i]^{neighbors}} \right|$

- The best candidate to replace the dead cell is the one with the lowest Score.

4.2.3 Cooperative Nursery

The cooperative nursery although related to the balance nursery is based on resource criticalities and favors an equilibrium between cell egoism and altruism. That is, the new cell not only needs to provide the resources required by the neighboring nodes (altruism) but also requires to find in that position the resources essential to its own survival (egoism). This cooperation between the environment (and indirectly the neighboring cells) and the new cell is what can ensure that the viability of the whole is maintained.

The algorithm of the cooperative nursery is as follows (see Algorithm 2 below):

- For each external resource ER_i, evaluate the highest and lowest criticalities on neighboring nodes.
  - $maxECrit_i = \max_{neighbors}(ECrit_i)$ (Altruistic behavior)
  - $minECrit_i = \min_{neighbors}(ECrit_i)$ (Egoistic behavior)
- Sort these criticalities into two sets {$maxECrit_i$} and {$minECrit_i$}.
- Select randomly one set and apply the associated process to select cell candidates. If no candidates can be found, remove the processed resource from the set, change the working set and repeat the process until some candidates are found. If all resources in both sets have been tested and there are still no candidates, the dead cell stays dead for this simulation cycle. The process applied at each iteration is as follows depending on the selected working set:
  - If {$maxECrit_i$} is selected the node applies an altruistic behavior: For the most critical resource in the set, filter out candidate cells that did not release this resource over a time window in the recent past. Only candidates that can help this neighborhood by producing and releasing this resource are selected.
    - If several candidates released the resource, keep only the top half candidates (in terms of resource quantity released).
  - If {$minECrit_i$} is selected the node applies an egoistic behavior: For the less critical resource in the set, filter out candidate cells that did not gather this resource over a time window in the recent past. Only candidates that can thrive in this neighborhood are selected.
    - If several candidates gathered this resource, keep only the top half gatherers (in terms of resource quantity gathered).
- If there are still several candidates after all resources have been tested, the environments of these candidates and of the dead cell are compared using a Manhattan distance. The candidate cell with the smallest distance is selected:
  - $Dist_{candidate} = \sum_i^{resources} \left| [ER_i]^{candidate} - [ER_i]^{dead} \right|$

Algorithm 2 Cooperative Nursery

{$maxECrit_i$} ← Sorted set of maximum resource criticalities on neighboring nodes
{$minECrit_i$} ← Sorted set of minimum resource criticalities on neighboring nodes
{$pCell_i$} ← Set of candidate cells around node that could divide (empty at the beginning)
{$ECrit_i$} ← Working set of resource criticalities: randomly initialized from {$maxECrit_i$} or {$minECrit_i$}
DO
    IF ({$ECrit_i$} = {$maxECrit_i$})
        Consider the most critical resource R_j in {$ECrit_i$}
        FOREACH (C_k candidate in {$pCell_i$})
            IF (C_k did not release R_j over time window)
                Remove C_k from {$pCell_i$}
            ENDIF
        ENDFOR
        IF (size of {$pCell_i$} > 1)
            {$pCell_i$} = top half producers of R_j
        ENDIF
        Remove R_j from {$maxECrit_i$}
    ELSE
        Consider the least critical resource R_j in {$ECrit_i$}
        FOREACH (C_k candidate in {$pCell_i$})
            IF (C_k did not gather R_j over time window)
                Remove C_k from {$pCell_i$}
            ENDIF
        ENDFOR
        IF (size of {$pCell_i$} > 1)
            {$pCell_i$} = top half gatherers of R_j
        ENDIF
        Remove R_j from {$minECrit_i$}
    ENDIF
    Switch {$ECrit_i$} between {$maxECrit_i$} and {$minECrit_i$}
WHILE (size of {$pCell_i$} < 1 and ({$minECrit_i$} != Ø or {$maxECrit_i$} != Ø))
IF (size of {$pCell_i$} > 1)
    Use Manhattan distance to select the best candidate
ENDIF
IF (size of {$pCell_i$} = 0)
    No new cell for this cycle
ENDIF

4.3 Reaching a Dynamical Equilibrium without Mutations

In this version of the simulation, cells do not mutate. That is, when a cell dies it is replaced by an exact copy of the selected neighboring cell. The choice of the copied cell is done according to the nurseries described in the previous section. The random nursery is used as a baseline to highlight the system behavior when no altruistic or cooperative decision is applied and to show that these nurseries behaves as expected by stabilizing the system.

In order to test the efficiency of the node agent decision when selecting a cell to replace a dead one, several scenarios are presented. The testing strategy is first to design a very simple system where the choice of a new cell has a strong impact on the survival potential of the system. Also, the system must be able to reach a dynamic equilibrium that is stable even when part of its population is regularly replaced.

The tested scenarios involve interdependent cell types. To live, a cell requires energy that is directly proportional to the amount of a single specific resource. It is able to produce this energy resource using two other resources. The interdependency is based on the fact that the energy resource produced by one type of cells is required by another different type of cells to produce their own energy resource (without the cells knowing this, of course). The difference between cell types is the resource used as energy supply and the production actions.

Considering that nCtype of cell types are present in the system, for the cell type $j \in \{1, \dots, nCtype\}$ the energy resource is $IR_{\left(j+\frac{nR}{2}\right)}$ from the set of rare resources, SetB. According to the basic actions presented in the previous chapter (see section 3.2.5.2), the actions associated with this cell type are defined as follows:

- Gather resource actions: For resource $i \in \{1, \dots, nR\}$, $gA_i: ER_i \xrightarrow{gA_i^{cost}} IR_i$
- Release resource actions: For resource $i \in \{1, \dots, nR\}$, $rA_i: IR_i \xrightarrow{rA_i^{cost}} ER_i$
- Cell type specific production action: Given the couple $(a_{j1}, a_{j2})\ with\ a_{j1} + a_{j2} = 1$, specific to each cell instance,

$$pA_j: a_{j1}.IR_j + a_{j2}.IR_{\left(j \bmod_{nCtype} + \frac{nR}{2} + 1\right)} \xrightarrow{pA_j^{cost}} IR_{\left((j+1) \bmod_{nCtype} + \frac{nR}{2} + 1\right)}$$

This set of equations ensures that each cell type is dependent on another cell type for its survival and that it is itself required for the survival of another cell type.

At the start of the simulation, each cell instance of a given type j has a unique random couple (a_{j1}, a_{j2}). This allows the system to cope with the lack of mutations by introducing enough diversity at the start of the simulation. Cells of a given type are spatially grouped together at the beginning of the simulation and the proportions for each cell type are the same (this can be seen, for example, in Figure 4-1 on the left picture, and in

Figure 4-3 on the right picture).

To thrive, any cell of one type must be at a certain distance from a cell of the other complementary type. If the decision mechanism works well, cells of these types should mix to optimize the exchange of critical resources. At the same time, once a dynamic equilibrium is reached, individual cells should be able to survive long enough so that cell death becomes a rare event.

The question is: Is it possible to balance a system where a single cell type that disappears represents the doom of the whole? On top of that, in order to achieve the goal, each cell needs to release its energy resource. That is, it must relinquish some of its lifespan for the community survival. Of course, in order to avoid any easy bias towards the use of the resources, a cell does not "know" which resource is vital for it.

For the results given below, the parameters of the simulated system are as follows: nCtype is set to 2 or 10, nR is set to 20 *i.e.* $SetA = \{R_0 \dots R_9\}$ and $SetB = \{R_{10} \dots R_{19}\}$. The radius used for candidate lookup in the nursery is 4 nodes. The system is initialized with a random distribution of

renewable abundant resources from SetA inside the cells and on the environment nodes. Rare resources from SetB are randomly distributed inside the cells to provide them with enough potential to survive 80 cycles (larger values do not strongly impact the outcome). In the environment, resources from SetB are set to 0.

Considering the scenario where there are nCtype is set to 2, the difficulty is that at least one of the resources to produce the energy of cell type 1 is the energy resource produced by cell type 2. Similarly, the energy resource of cell type 1 is a required component of energy production in type 2 cells. This interdependency between cell types creates the basic layout for future division of labor and the emergence of communication protocols.

As mentioned earlier there are multiple possible criteria for the node agent to decide which cell has to be duplicated on it when a cell dies. The following node agent nurseries have also been tested with the setup previously presented without leading to any interesting results:

- Age: Cells that are old are by definition adapted to their environment. Selecting the oldest cell looks like a possible strategy. Unfortunately, a cell that is able to thrive in one place and a given local environment is not necessarily able to survive in a different environment.
- Age*Energy: This variation takes into account the potential future of the cell given its energy stock. Although this works better than age alone, long term stability cannot be reached.
- Resource production: Cells that produce the most could be the interesting ones since they contribute to the common goods in the system. But maximum production can also result in a short lifespan because of bad regulation mechanisms.
- Environment similarity: This assumes that if environment changes are the same over a period of time then the cell can live on its node of origin or on the dead cell node. So, the variations of resources on a time window are compared between the candidate cell node and the dead cell node. This criterion only looks at the environment modifications and does not take into account the cell. This does not work well because it is not guaranteed that the new cell can find what it needs for its own survival. Although completely unbiased, this algorithm cannot help reach or maintain a dynamic equilibrium.

Since none of these nurseries successfully maintained a steady state of the system and none led to any interesting results, they will not be discussed any further.

4.3.1 2 Interdependent Cell Types, 20 Resources

The size of the system is 60×60 (Figure 4-1). The production rules for the two cell types are:

$$\text{Cell type 1: } a_{11}.IR_1 + a_{12}.IR_{10} \rightarrow IR_{11} \; and \; a_{11} + a_{12} = 1$$

$$\text{Cell type 2: } a_{21}.IR_2 + a_{22}.IR_{11} \rightarrow IR_{10} \; and \; a_{21} + a_{22} = 1$$

Cell type 1 gets its energy from R_{11} and cell type 2 from R_{10}. Every cycle a cell has to choose between 20 gather actions, 20 release actions and 1 production action.

The initial conditions and parameters of the simulation are set in a way that is quite demanding on the system. For example, there are no SetB resources present in the environment. This means that cells need very quickly to be able to produce and release R_{10} and R_{11} for the interdependence production rules to work and the system to survive as a whole. Also, the rate of disappearance for these resources is set high enough in order to avoid accumulation in the system. Too much resources renders the selection process inefficient since any kind of cell can survive in these conditions. Cell type distribution is another parameter that can put strains on the system. When distributed randomly on the grid, cells of each type have a good chance to have several cells of the

other type around them. This would mean an easy access to the required resource for their energy production. To make it more difficult and stress the nursery process, cells of each type are grouped together at initialization time. So, the dynamics of the dead cell replacement mechanism has to mix them in order to provide the needed resources to all cells.

When this system setup is simulated using the altruistic nursery (section 4.2.2) a dynamic equilibrium is found and patterns of cell distribution emerge as depicted in Figure 4-1. Interleaving of the two cell types, as observed at steady state is an efficient way of exchanging the critical resources before they are removed from the system.

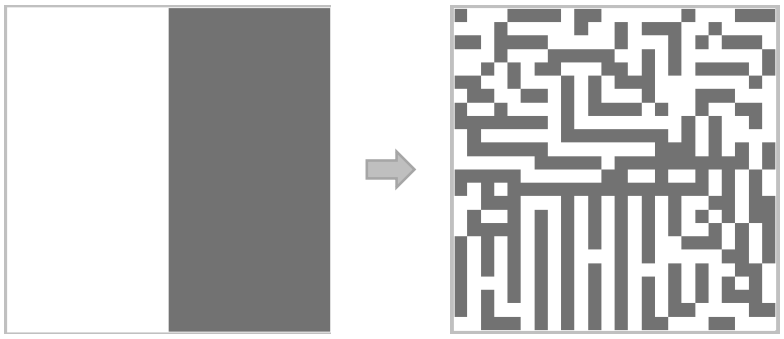


Figure 4-1 Cell types distribution pattern observed using the altruist nursery

Even for large systems of 100×100 cells, a dynamic equilibrium can be reached and maintained. But this algorithm reaches its limits with systems containing more than two interdependent cell types. Starting with three cell types, the balance nursery is unable to select cell replacements that ensure long term stability of the system. One cell type always disappears after several thousand cycles causing a death cascade of other cell types and the final system demise.

Although a promising approach, simply trying to balance the inputs and outputs is not sufficient to ensure the stability of large and diverse systems. This is probably because the needs of the new cell might not be matched by what its destination node can provide.

Using the cooperative nursery, the results are somehow similar since the system is also able to reach a dynamical equilibrium but the details of the evolution of the system are different and the process is more robust since it can deal with more than two cell types as described in section 4.3.2. the cell type distribution is presented in Figure 4-2.

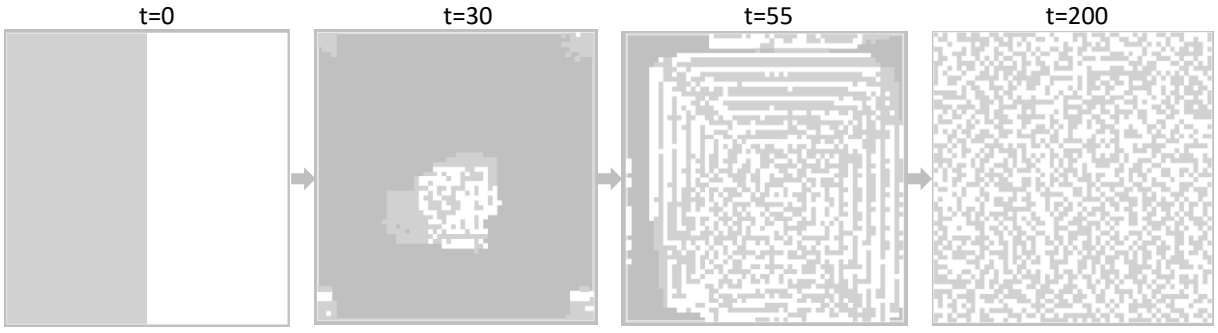


Figure 4-2 Evolution of a system with 2 cell types and 20 resources. Green dots are cell type 1 and white dots cell type 2. Dead cells are in dark grey.

The left and center graphs in Figure 4-3 represent the evolution with time of the average concentration of resources inside the cells of the system. On the left graph, a short transition phase of less than 1000 cycles takes place where concentrations of resources decrease quickly from the initial random distribution to levels where criticality starts to play a role. It can also be observed

that resources R_{12} to R_{19} completely disappear from the system after 300 cycles since none of the two cell types can produce or use them. These resources are then always the most critical ones but no action can improve their state. This does not endanger the system as a whole since these resources do not impact the viability of cells.

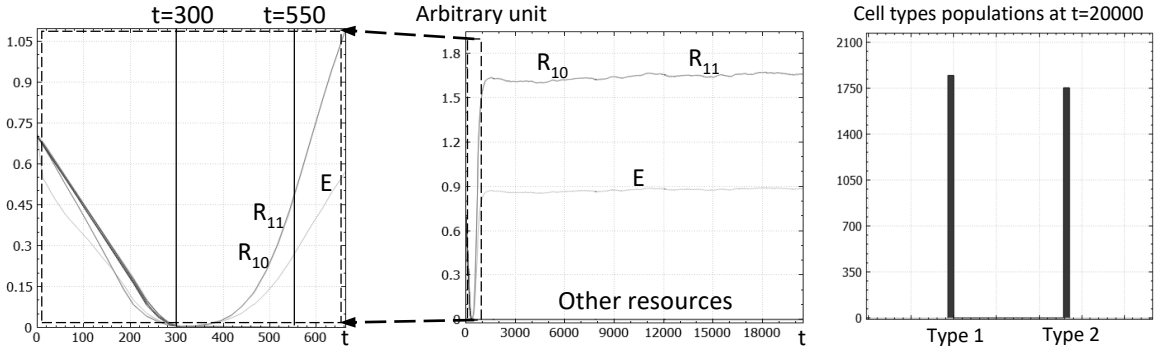
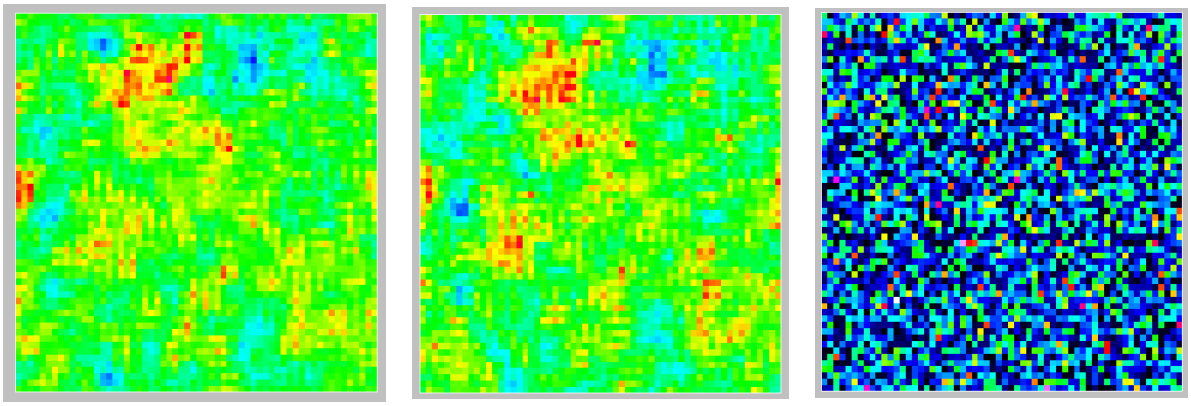


Figure 4-3 Left and center: Energy and average concentration of resources in the system. The graph on the left is a close-up of the first 650 steps shown on the graph in the center. Right: Number of cells of each type at the end of the simulation showing that an equilibrium between cell populations has been reached.

In contrast, resources R10 and R11 drop for the first 300 cycles (Figure 4-3, graph on the left) until they become critical for the survival of the system. During this time, cells die and are replaced by more efficient instances by the nodes. At t=300, less than 10% of cells are still present and if the nursery process is not efficient there might not be enough cells of each type to reseed the system (Figure 4-2). When cell death is not too sudden and cells are well mixed, between t=300 and t=550, there is a growth phase where instances of each cell type divide to replace all dead cells and refill the empty spaces (Figure 4-2, right). Then we start to observe the effective regulation of the essential resources (R10 and R11) and their curves raise slowly to an optimal level where the dynamical equilibrium is reached and maintained. The actual convergence values (about 1.6 in Figure 4-3 middle graph) for resources are emergent properties of the system since there is nothing in the rules and decision process that is related to this.

Figure 4-4 shows the distribution of resources R_{10} and R_{11} at the equilibrium (t=20000). They are well distributed and no part of the system dangerously lacks any of them. Similarly, the age distribution does not show any noticeable region, indicating that the system is homogenous and stable.



R_{10}: min_{blue}=1.48, Max_{red}=1.71 R_{11}: min_{blue}=1.45, Max_{red}=1.74 Age: min_{black}=1, Max_{red}=1308

Figure 4-4 t=20000: Resource distribution of R_{10} and R_{11} in the system and age distribution

It is noteworthy to mention that the spatial distribution of type 1 and type 2 cells is not regular and does not display any kind of visually distinctive pattern (Figure 4-2 last image) as observed with the altruist nursery.

In the case where random selection of new cells is used, the system behaves similarly for the first thousand cycles but then is unable to maintain a dynamic equilibrium of rare resources from SetB everywhere in the grid. Figure 4-5 shows the evolution of resource R_{10} and the distribution of cell age or cell type with time. Heterogeneities start to appear in the distribution of resources and soon this leads to local cell death. The random nursery does not always place cells that could correct the lack of resources and therefore the unbalance only grows with time creating larger and larger regions of dead cells. After about 11k cycles the last cells die and the system is restarted with a different random configuration (Figure 4-6). Surprisingly enough the cell type population distribution stays balanced until the end. But this is probably because random choices between two species should always be 50% if the random generator is good enough and that neither cell type has a survival advantage over the other.

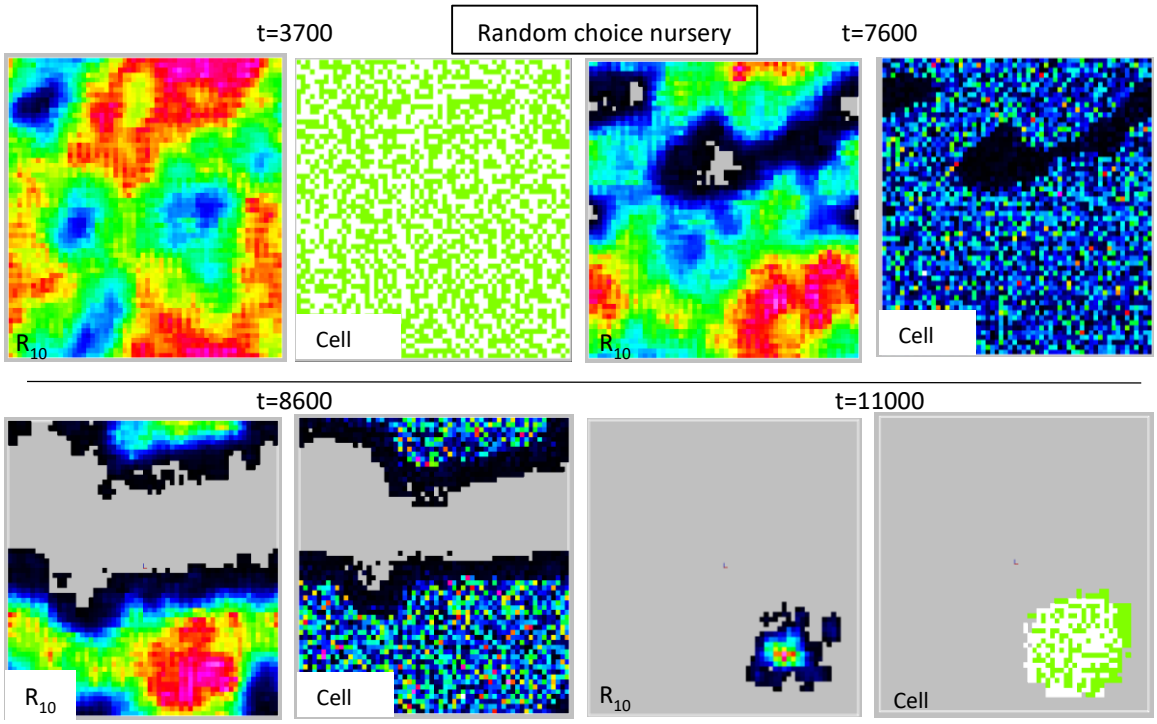


Figure 4-5 System evolution using the random choice nursery

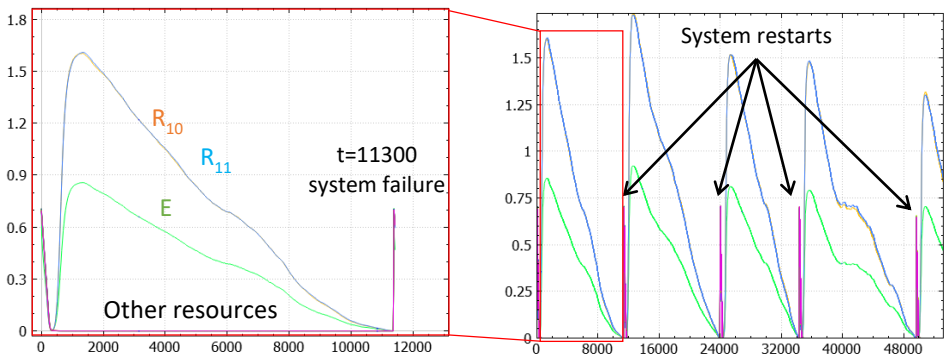


Figure 4-6 Energy and average concentration of resources using a random choice nursery. The peak at 1130 represents a system restart

Interestingly, for the cooperative nursery, after a few thousand cycles, only a few instances of type 1 and type 2 cells are left from the initial random sets and occupy all the simulation space. The surviving instances selected by the cooperation process always have production factors a_{j1}[8] above 0.8 and a_{j2} below 0.2. This can be explained by the fact that cells using more scarce resources from SetB will have a shorter lifespan and be replaced more often. This also indicates that the system maximizes the use of abundant resources (ER_0 to ER_9) and minimizes the consumption of rare resources (ER_{10} and ER_{11}). To understand how and when this selection happens the distribution of a_{j1} is analyzed in Figure 4-7. Starting from a (quasi)uniform distribution of values between 0.05 and 0.95 the system quickly favors higher values as shown at t=380. Although only about 5% of the cells are alive at this stage there are still some low a_{j1} values present. During the expansion phase that follows, further selection occurs and only a_{j1} values higher than 0.9 remain at t=650. From there the system reaches a dynamic equilibrium where no further evolution can happen for lack of diversity.

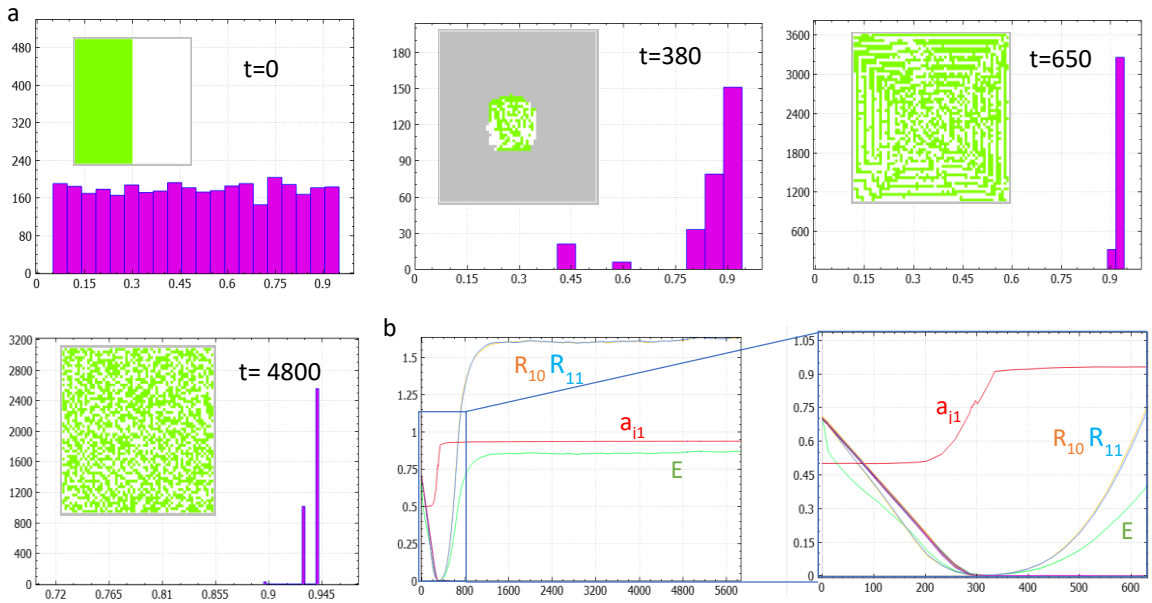


Figure 4-7 Evolution of a_{j1} during the simulation. a) Distribution of a_{j1} at different times with the corresponding cell types populations. b) Evolution of resources and average a_{j1} with time

To understand if the instance selection is influenced by the system starting conditions, a simulation is performed using less harsh initial conditions and parameters. In particular, rare resources from SetB are distributed in the environment (1.6 instead of 0) so that the system resembles the final stable state. In this way, cells should have enough resources and a_{j1} selection should be less necessary. The result is that all graphs are similar to Figure 4-7 and even in these conditions SetB resources decrease (although more slowly) to a point where cells die and the factor selection occurs. Higher starting values do not remove this phase, they only delay it.

These results show that the system inevitably converges to a population that is less dependent on the rare resources and also that cells with low a_{j1} jeopardize the potential to equilibrate the system as long as they are present, independently of the starting conditions.

A question raised by this experiment is whether the nursery has any role in this process or not? The same analysis performed on a simulation using a random nursery leads to the same conclusions, that is a factor selection phase exists. This would mean that the cooperative nursery

[8] This factor measures the ratio of abundant resources versus rare resources used in the production action. High values indicate a preference for the consumption of abundant resources.

does not impact this selection process and does not introduce any bias (at least for that phenomenon).

4.3.2 10 Interdependent Cell Types, 20 Resources

In this experiment, 10 interdependent cell types are present in a 60×60 system and regulate 20 resources, 10 of which are flowing out of the environment at a constant speed (SetB). The production rules for the various cell types are:

Cell type 1: $a_{11}.IR_1 + a_{12}.IR_{10} \rightarrow IR_{11}\ and\ a_{11} + a_{12} = 1$

Cell type 2: $a_{21}.IR_2 + a_{22}.IR_{11} \rightarrow IR_{12}\ and\ a_{21} + a_{22} = 1$

Cell type 3: $a_{31}.IR_3 + a_{32}.IR_{12} \rightarrow IR_{13}\ and\ a_{31} + a_{32} = 1$

Cell type 4: $a_{41}.IR_4 + a_{42}.IR_{13} \rightarrow IR_{14}\ and\ a_{41} + a_{42} = 1$

Cell type 5: $a_{51}.IR_5 + a_{52}.IR_{14} \rightarrow IR_{15}\ and\ a_{51} + a_{52} = 1$

Cell type 6: $a_{61}.IR_6 + a_{62}.IR_{15} \rightarrow IR_{16}\ and\ a_{61} + a_{62} = 1$

Cell type 7: $a_{71}.IR_7 + a_{72}.IR_{16} \rightarrow IR_{17}\ and\ a_{71} + a_{72} = 1$

Cell type 8: $a_{81}.IR_8 + a_{82}.IR_{17} \rightarrow IR_{18}\ and\ a_{81} + a_{82} = 1$

Cell type 9: $a_{91}.IR_9 + a_{92}.IR_{18} \rightarrow IR_{19}\ and\ a_{91} + a_{92} = 1$

Cell type 10: $a_{101}.IR_{10} + a_{102}.IR_{19} \rightarrow IR_{10}\ and\ a_{101} + a_{102} = 1$

All resources are necessary and need to be regulated in order for the system to survive.

This experiment is much more challenging than the one using 2 cell types. Since the resources from SetB cannot diffuse far before being removed from the system, each cell must be close enough from both a cell that provides the resource needed to produce its energy resource and a cell that depends on its own production. If a single cell type completely disappears from the grid, it is the whole system that is doomed to fail. Therefore, starting conditions and parameters similar to the ones used in the "2 interdependent cell types" experiment do not work in this context. In particular, the first "mass extinction" phase observed in this previous experiment at around 300 cycles, happens now in around 500 cycles. It is not possible to find conditions where 10 cell types are mixed enough to survive this mass extinction in such a short time. Thus, conditions in this experiment have to be somewhat less constraining. In particular, to soften the starting conditions, the energy expenditure per cycle and SetB resource removal are reduced. In this way rare resources can diffuse farther from their emission point and can be used by more cells to survive. This results in a slower simulation pace but is compatible with the number of cell types to deal with.

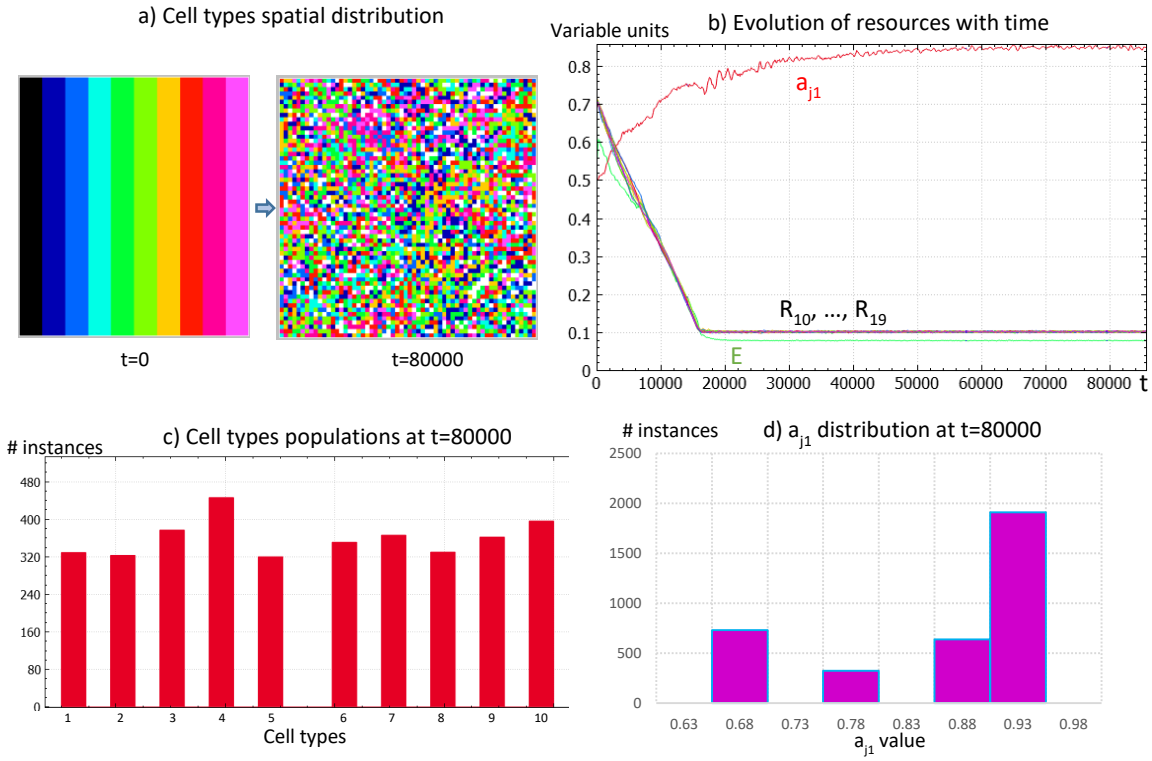


Figure 4-8 Simulation of a 10 interdependent cell types with 20 resources in a 60×60 grid

Given these gentler conditions, it is not surprising to observe a much slower decrease in resources in the first stage of the simulation where cell types mix together (Figure 4-8 b). After around 16000 cycles, global resources have decreased by a factor of 7 but, during this time, a_{j1} selection occurred and the system becomes performant enough to start regulating efficiently all resources from SetB. Since the number of cell instances per cell type is 5 times less than in the "2 interdependent cell types" experiment, the diversity in a_{j1} values is consequently restricted. This explains why even when the dynamic equilibrium is reached, there are still many cells using a_{j1} factors below 0.8 (Figure 4-8 d). Actually, it can happen that all the instances of a particular cell type are stuck with low a_{j1} because after a few thousand cycles the diversity was gone from the system for this cell type. The fact that high a_{j1} disappeared instead of low a_{j1} can be circumstantial and linked to the limited size of the system. For example, this could happen when the last cell with a_{j1}=0.95 is far from any cell providing its critical resources while several cells with a_{j1}=0.3 are situated in a more suitable neighborhood. Quickly, only cells with a_{j1}=0.3 are present everywhere in the system.

In this experiment, the number of cells per type is well balanced and differs on average by 8.2% and at most by 16% from the ideal value of 360 (Figure 4-8 c). This number decreases when the size of the system increases. This means that random noise contributes for a significant part to this value in smaller systems. For example, in a 200×200 simulation (Figure 4-9) this deviation becomes 2.3% and is at most of 4.6%.

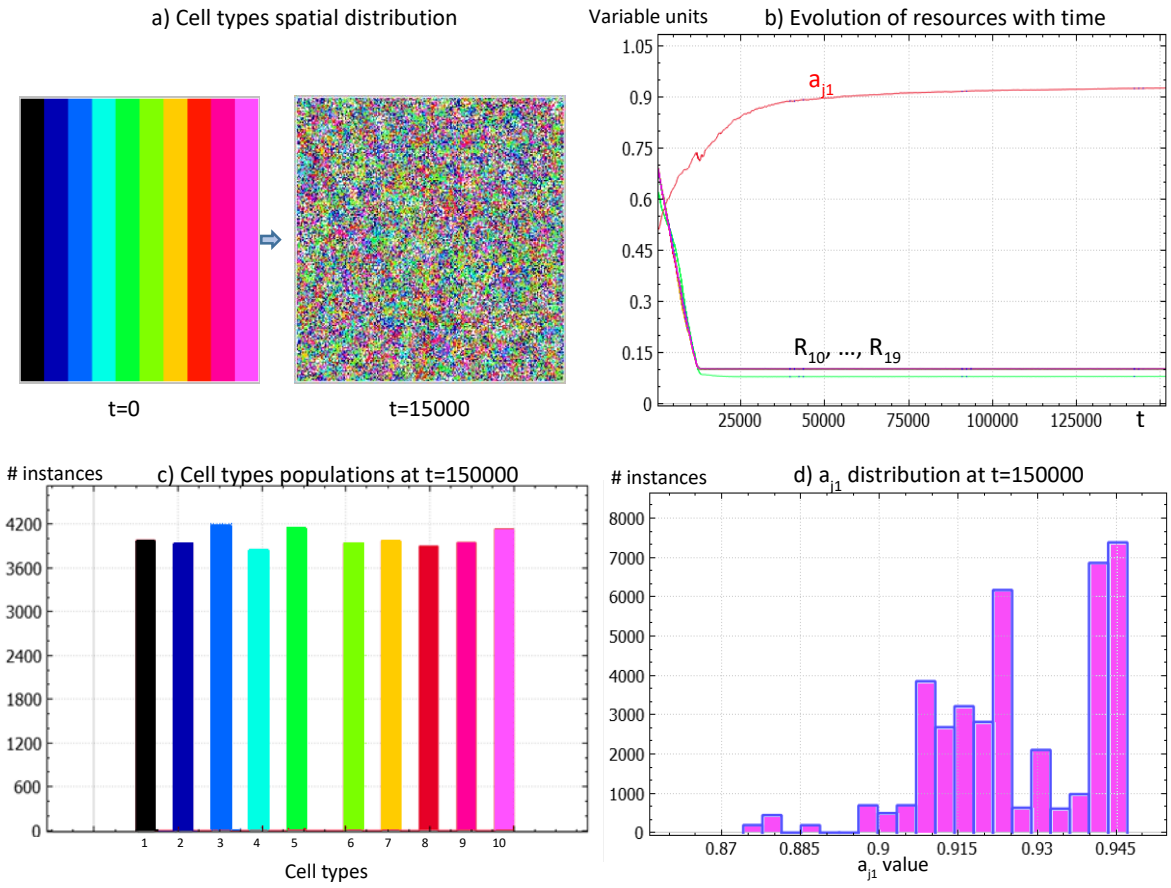


Figure 4-9 Simulation of a 10 interdependent cell types with 20 resources in a 200×200 grid

The results obtained with these experiments demonstrate the ability of the cooperative nursery to efficiently regulate a group of interdependent cell types. In the following sections we investigate various properties of this cooperative nursery.

4.3.2.1 Solution Stability

In the last experiment, after 50000 cycles the system has a balanced cell types population distribution and high a_{j1} values that ensure a near optimal use of resources. Some 100000 cycles later, the dynamic equilibrium still exists and it could be assumed that in these conditions the system can persist as it is by itself even using a random choice for new cells.

Therefore, a new experiment was performed to investigate the stability of the system if the cooperative nursery is switched to a random nursery once the system is in a steady state as represented in Figure 4-10 where such a transition is done at around 55k cycles. The result shows

that although the system is able to survive for longer than in a pure random selection system (20k cycles vs 10k cycles) it eventually fails.

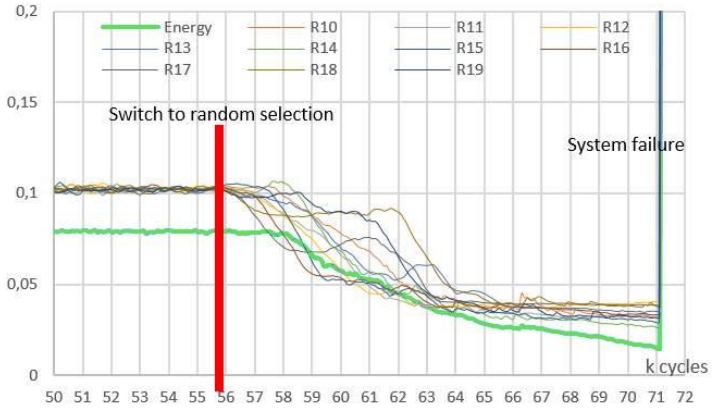


Figure 4-10 Switch from cooperative nursery to random nursery

Intuitively, this should not be the case since once a natural system has reached a dynamic equilibrium in a given environment and if conditions do not change it should survive.

This apparent contradiction can have several explanations. For one, a natural system is usually huge compared to our simulation setup, and if a part of such a system fails, it can be recolonized by populations from other locations. Secondly, a natural system cell division process is not pure randomness, there are rules that dictate which cell can divide, when and where (H. Lodish, Berk, and Zipursky 2000). A comparison of the cooperative nursery should objectively be performed against this natural nursery. However, we do not fully understand these rules and they would probably not apply in our simplistic system.

4.3.2.2 Robustness to Noise

An interesting property to study is the resilience to noise of the cooperative nursery. That is, to what extent new cells may be chosen randomly in the neighborhood of the dead cell, without the system collapsing?

To illustrate this, the following experiment is performed: Starting with the same system setup as in 4.3.2 (10 interdependent cell types), simulations are run using a probability ranging from 10% to 100% of using the random nursery instead of the cooperative one. For each percentage, the stability of the system is assessed for 200k cycles.

Figure 4-11 presents results for some of the percentages. The curves represent the average energy of cells in the system with time. Each peak indicates that at least one cell type disappeared from the grid resulting in a system restart. We can observe that 100% random nursery is always unstable, as mentioned before. In all other cases, the system is fairly stable and failures do not seem to depend on the random percentage since three restarts are present at 10%, four at 96% and zero at 99%. Several simulations with the same initial parameters and conditions were performed and gave the same average number of restarts (0-4).

This therefore seems representative of the behavior of the system, and random initial conditions seem to be the main factor of the number of failures. This means that the system is more sensitive to initial conditions when the cooperative nursery is not always used, but that its usage frequency does not have a strong impact. In other words, the system is binary since without cooperative

nursery it is always unstable but with it, stabilization is possible irrespective of the percentage of randomness.

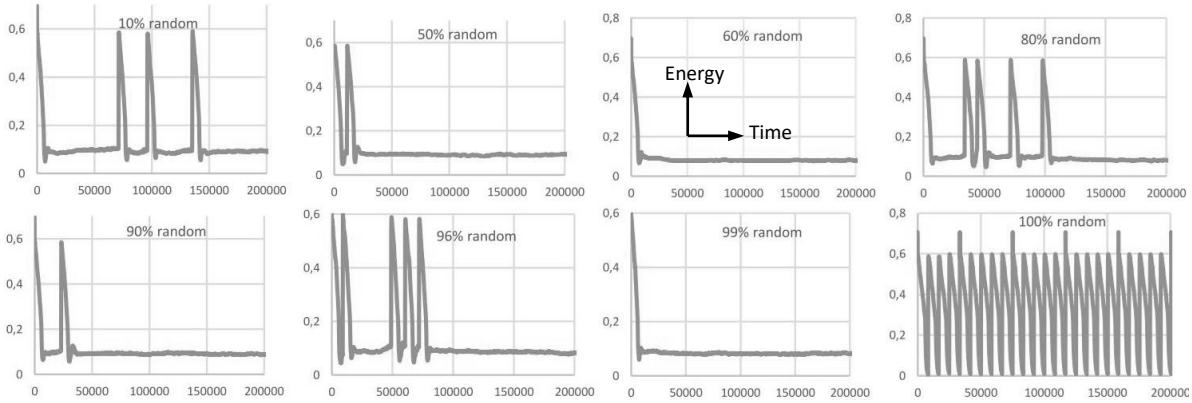


Figure 4-11 Stability of a 10 cell types system with increasing percentage of random nursery. Curves represent the average energy and peaks represent system failures

4.3.2.3 Saving a Failing System

The next property studied was the capacity of the cooperative nursery to rescue and regulate a system that is on the verge of collapse. To do so, an experiment similar to 4.3.2 was performed using a random nursery until the system fails around 23000 cycles. Every thousand cycles the system is stored. Then all these stored systems are independently resumed using the cooperative nursery. Finally, the survival time of these systems is measured. It is important to note that this is not equivalent to using different random initial conditions since in the stored systems the a_{j1} values, the cell type distribution and the cell resources are constrained by the evolution of the simulation so far.

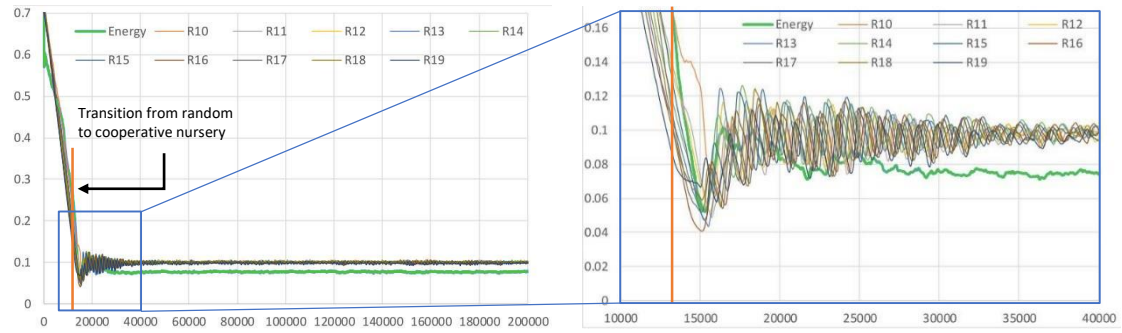


Figure 4-12 Switch from random to cooperative nursery

Figure 4-12 shows that the cooperative nursery is able to rebalance the system with the random nursery prior 15000 cycles of simulation. After that blurred frontier, the system cannot be saved. This is usually because one cell type is too close to extinction to reproduce before old age or lack of energy completely remove it from the simulation. Another reason is that the a_{j1} distribution still present in the system is not high enough for a balanced system to persist.

It is interesting to note that after the switch, the system goes through a transition phase where resources oscillate (see right part of Figure 4-12) as redistribution of cell types on the grid tries to rebalance the system.

4.3.2.4 Scalability

Some nurseries other than cooperative are able to balance particular systems. As presented in 4.2.2 the balance nursery can deal with a system composed of two cell types but fails with more diverse ones. In order to stress the cooperative nursery capacity to regulate large systems, a simulation of 300×300 with 30 interdependent cell types and 60 resources (30 rare ones) is tested. Although much longer to run, the system is stable after 100 cycles (Figure 4-13). This suggests that the regulation process scales well with all parameters of the system.

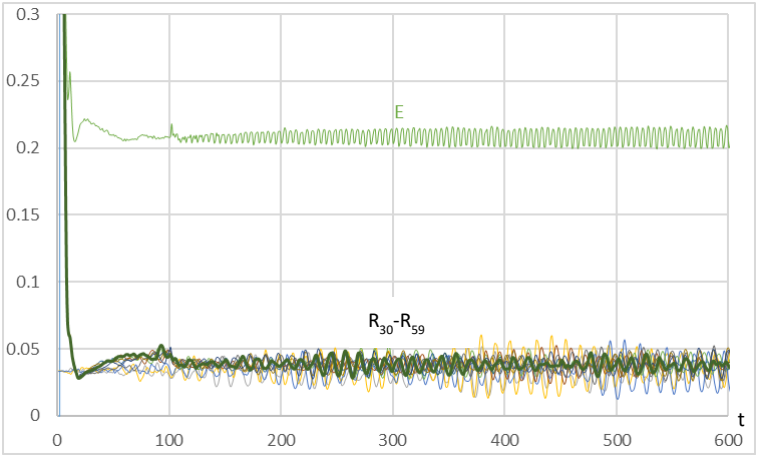


Figure 4-13 Simulation of a 300×300 grid system with 30 cell types and 60 resources

4.4 Conclusion on the Cooperative Systems without Mutations

In this chapter, we have implemented the AMAS model proposed in chapter 3 to simulate a multicellular tissue composed of very simple interdependent cells. Two important cooperative behaviors have been described, one for the decision-making of a cell agent and the other for the dead cell replacement processed by a node agent.

As we have demonstrated with the different experiments presented in this chapter, various node agent nursery behaviors can be implemented to decide mother cell candidates. Most of them are unable to maintain a dynamic equilibrium in a simple system composed of two interdependent cell types and only one succeeded in regulating more complicated systems. The cooperative nursery showed its potential to accelerate the convergence of various systems towards a stable steady state. Furthermore, it is able to withstand some measure of noise before failing to regulate the system. Finally, its performance scales well with the size of the system or the number of cell types and resources.

The cooperative attitude of cell and node agents when performing respectively action selection or replacement of a dead cell is a means to accelerate the convergence of the simulation. But it must be done without introducing any bias that could favor a particular final state of the system. The true drive of the simulation (and of a real system) is cell survival and this is mainly linked to the capacity of cells to keep their energy levels above zero and to function properly. As described, the production factor (a_{j1}) selection process that naturally occurs in favor of energy saving cells is not altered by the cooperative nursery which is a good indicator that if there is any bias, it is secondary to the energy drive that is still controlling the fate of the system. Furthermore, to stabilize a system, the cooperative nursery does not need to be used constantly as shown in 4.3.2.2. Thus, if needed, the system could be simulated using 99% of random selection of new cell to be closer to natural systems. In that case the bias introduced would be minimal although still present (if it exists).

These results represent a good base to further develop a simulation to observe the emergence of communication in an unbiased fashion.

In the next chapter, mutations are introduced in the system in order to allow cells to actively adapt to their environment. This is also a necessary process for the emergence of new behaviors like the exchange of information between the cells. Indeed, the specialization of a resource into a signal can only happen through cell mutations since nothing in the system distinguishes one resource from another at the beginning. Modification of cell actions is thus necessary in order to focus the cell role on specific productions and conditional release of resources which are a prerequisite for communication.

CHAPTER 5. COCELL2: EVOLUTION THROUGH MUTATIONS

Dans ce chapitre, les mutations sont mises en œuvre comme deuxième étape de la simulation. Nous savons de CoCell1 qu'un système avec de fortes interdépendances et de nombreux types de cellules peut atteindre un équilibre dynamique si les bonnes actions de production sont présentes dans les cellules et si la distribution des types de cellules permet un flux de ressources efficace.

Dans CoCell2, des mutations sont introduites pour adapter une nouvelle cellule à son environnement. Ce processus, s'il est effectué correctement, devrait améliorer la vitesse et la convergence du système vers un équilibre dynamique. Il peut aussi laisser émerger des comportements globaux qui ne sont pas codés dans les actions cellulaires comme la communication cellule-cellule locale.

In this chapter, mutations are implemented as the second stage of the simulation. We know from CoCell1 that a system with strong interdependencies and many cell types can reach a dynamic equilibrium if the right production actions are present in the cells and the distribution of cell types allows an efficient resource flow.

Here, mutations are introduced to adapt a newborn cell to its environment. This process, if done correctly, should improve the speed and the convergence of the system towards a dynamic equilibrium. It can also let emerge global behaviors that are not encoded in the cell actions like local cell-cell communication.

5.1 Cooperative Mutations

CoCell1 can reach and maintain a dynamic equilibrium in a constraining but dynamically constant environment. If conditions change drastically once the system is stable, it inevitably fails. This is because convergence and optimization of the cell parameters focused the system on high performance cell instances that work well in a given environment. Other instances with different parameters are discarded during this process and adaptability suffers. To be able to adapt to a changing environment, diversity needs to be reintroduced in the system and this is the role of mutations. Furthermore, as mentioned in section 2.7.2 mutations are unavoidable in natural systems so they need to be integrated into the simulation.

Mutations in our system cannot be random as evolution would take too much time. The process needs to change cell actions in a way that directly improves its fitness[9] to the environment. As with

[9] Fitness here refers to ecological fitness and not to the fitness used in genetic algorithms or neural networks, although they can be related.

the decision process and the cooperative nursery, mutations selection can only use local information and avoid any kind of bias other than improved cooperation.

Although mutations can occur anytime during a cell lifespan for real organisms, in CoCell2 we decided to apply them only on newborn cells at the end of the nursery process. This allows for an easier control of this mechanism and is also the perfect time to adjust the new cell to a different environment since it is not at the same location as its mother. Also, we apply the mutation only to the new cell and not to its mother. Doing so, we do not perturb the environment of the mother cell and this should speed up the convergence of the system. It could be argued that mutations during the life of the cell could actually increase its chances of survival. This is probably true but will need another iteration of CoCell to implement and test.

Since production actions are the way a cell can impact its environment by its decisions, to be effective, mutations need to operate on them. So the first step is to design a generic cell type with a set of actions.

Unlike in previous simulations where each cell type energy source is different, in CoCell2 all cells share a common energy resource eR, without, of course, knowing this resource represents energy. This energy is produced from two rare resources R_a and R_b from SetB.

To avoid "cheating" by direct energy resource exchange between cells, eR cannot be gathered or released in the environment. Therefore, the generic cell has as many gather actions as there are resources minus one (nR-1) and as many release actions as there are resources minus one.

This generic cell has also a set of production actions (between 4 and 10 in the following examples). The first production action is responsible for the synthesis of the energy resource eR from R_a and R_b. All other production actions use two random resources, one from SetA and one from SetB to produce a random resource from SetB except eR. Thus, some cells are able to produce R_a, R_b, both or none of these essential resources.

5.1.1 Inducing Interdependence in a Mutation Context: Division of Labor

In the previous experiments interdependency is included by design in the cell "genome[10]": To produce its energy resource, one resource from SetA and one resource from SetB produced by another cell type are necessary. With mutations enabled, it is now possible to start from an undifferentiated set of similar cells and create the conditions to induce interdependency.

In real cells, division of labor is a good way to achieve better fitness and efficiency (Ispolatov, Ackermann, and Doebeli 2012; Rueffler, Hermisson, and Wagner 2011). For example, let us suppose that in order to survive, real cells must produce two resources. The chemical reactions associated with these productions can have very limited overlapping conditions. It could be that one of them requires acidic conditions to perform well and the second more neutral or basic conditions. Or both reactions might compete for the same rare catalytic enzyme or reactant. In those cases, both reactions can take place in the same cell but their respective yields will be low or their cost very high because specific compartments need to be maintained to isolate reactions. An alternative is to create two cell types from a common ancestor. One specialized cell type only performs the first reaction and can therefore maintain an acidic medium for optimal performances while the second cell type fulfils the second necessary reaction in more neutral conditions. In this case, both cell types are able to perform better than their ancestor. The caveat is that specialized cells require a means to transfer part of their production to other cells for the organism to survive. Either the cells just dump their production in the environment and hope to find what they require to be present

[10] Genome here refers to the ensemble of actions a cell can perform and is not related to genetic algorithms.

there or coordination via communication is required to optimize the whole process. A similar setup can be implemented in CoCell2 in the way described thereafter.

Since all cells produce eR from R_a and R_b, interdependency can arise if each cell produces either R_a or R_b but not both. In order to enforce this behavior, the energy cost of producing both resources in the same cell is multiplied by a constant factor. In this way, cells that produce both resources are independent and can exist but short lived. Cells that produce either resources are interdependent on cells that produce the complementary resource but are more frugal thus more likely to live longer if they can share with other cells. On the long term, survival selection is supposed to favor interdependent sharing cells in the system.

It is also possible that a cell does not produce neither R_a nor R_b. If the environment contains a lot of these resources, this kind of cell can survive by becoming what is commonly called a cheater. It means that it does not contribute to the common goods but thrive on them. In the long run these cells can become a threat by reducing the fitness of the system as a whole. In a competitive environment where other organisms compete for the same resources, this can result in extinction. In the simulation, appearance of cheaters is largely countered by a stringent environment where essential resources are depleted quickly by their transport outside the system.

A simple experiment can be performed to bring to light this phenomenon of division of labor. A single cell type (C_3) depends on eR for its energy. eR is produced from R_a and R_b, two rare resources from SetB. C_3 is able to produce both R_a and R_b from abundant resources from SetA. But the energetic cost for each of these production behaviors is multiplied by 20 when they are present in the same cell like C_3. Apart from these three production actions, each cell is also able to produce five different rare resources from abundant ones.

$$pAe_1: \quad a_1.\boldsymbol{IR_a} + a_2.\boldsymbol{IR_b} \rightarrow \mathrm{e}R \; and \; a_1 + a_2 = 1$$

$$pAe_2: \quad b_1.IR_{\mathrm{i} \in \mathrm{SetA}} + b_2.IR_{\mathrm{j} \in \mathrm{SetA}} \xrightarrow{Ecost_a} \boldsymbol{IR_a} \; and \; b_1 + b_2 = 1$$

$$pAe_3: \quad c_1.IR_{\mathrm{k} \in \mathrm{SetA}} + c_2.IR_{\mathrm{l} \in \mathrm{SetA}} \xrightarrow{Ecost_b} \boldsymbol{IR_b} \; and \; c_1 + c_2 = 1$$

If pAe_2 and pAe_3 are present in the same cell: $Ecost_a$ and $Ecost_b$ *20

The mutation process used in this example is copied from Nature: Randomness. Every time a cell dies, the cooperative nursery selects its successor and a mutation can happen with a probability inversely proportional to the age of the parent. The mutation when it occurs replaces one of the production actions with a random one producing a rare resource.

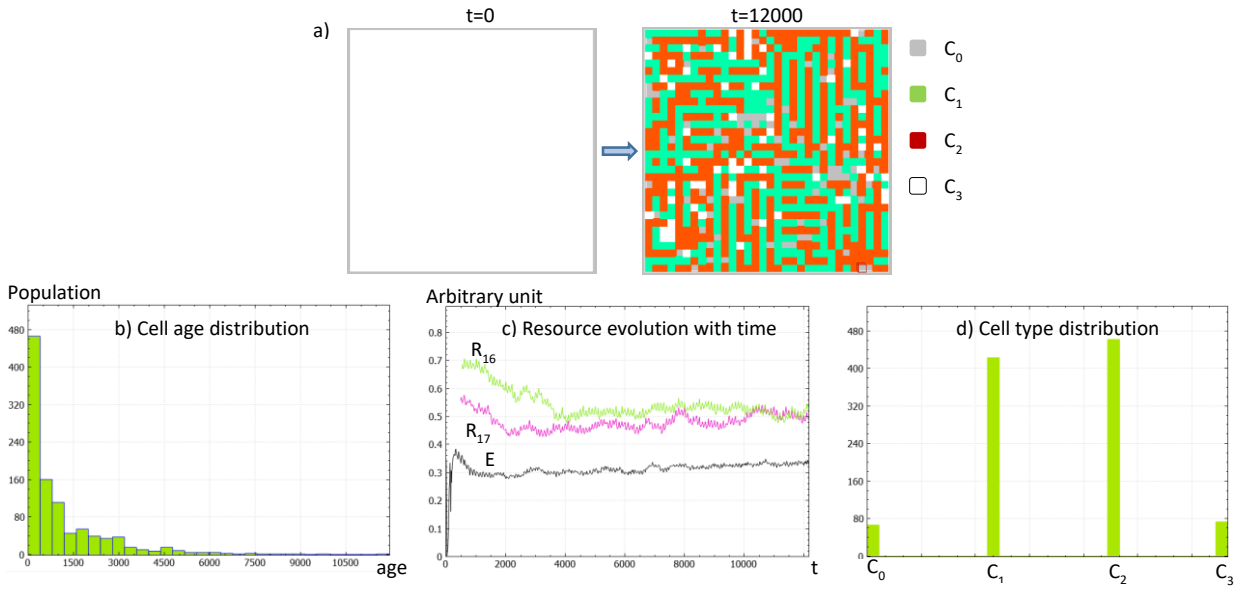


Figure 5-1 Division of labor induced by energy cost

Figure 5-1 presents the results of such a system. At the beginning of the simulation (t=0) all cells are of the type C_3 (white). As they consume a lot of energy to synthesize R_a and R_b they die quite quickly. By random mutations, C_1 and C_2 cell types appear that respectively produce R_a and R_b. Since producing only one of these resources has a normal energy cost, these cell types live longer and thrive. After a few thousand cycles, concentrations of R_a and R_b reach a dynamic equilibrium in the environment as it is regulated by C_1 and C_2 (Figure 5-1 c). The distribution of cell types at t=12000 presents a pattern of alternating C_1 and C_2 types (green and red) (Figure 5-1 a). The age distribution (Figure 5-1 b) shows that cells are able to survive much longer than the initial allocated energy resource and the cell type distribution indicates that C_1 and C_2 populations are well equilibrated. There are still some C_3 present that popups in the system from time to time due to mutations but as before they are short lived (Figure 5-1 d). There is also another type of cells that appears in the system: C_0 (gray) that produces neither R_a nor R_b. These cells are "cheaters" that survive by using the common goods present in the environment. Their proportion is similar to C_3 but they can actually live much longer since they do not spend energy to produce R_a or R_b. Actually in our setup even if they do not synthetize vital resources they nonetheless spend their energy on other non-essential materials. So their lifespan is equivalent to C_1 or C_2. Also, they are unlikely to become a large portion of the population since the cooperative nursery does not select them often. In real life, these cheaters can be much more conservative and preserve their own resources. In this case, the system as a whole might be in danger if they proliferate too fast.

This experiment demonstrates that division of labor and interdependency can be achieved and that cooperative decision-making and cooperative nursery do not interfere with this phenomenon. The mutation process must not alter this process either. In the following section, the mutation algorithm is described and its efficiency tested.

5.1.2 Orient Mutations to Speed up Evolution

In order to speed up evolution, all mutations need to have a positive impact on the fitness of a cell in its environment, and any negative/neutral mutation has to be avoided. Also, in order to somehow let the system stabilize in a good configuration, new cells are not always mutated. For example, the older is the mother cell (meaning it is quite adapted to its environment) the less

probable the daughter cell is mutated. This is done by using a probability of mutation that can be affected by various factors. It could be argued that detrimental mutations also play a role in evolution since they can end a branch of organisms that are in a dead end or take too much "space" in the ecological spectrum. Once these species disappear there is room for other organisms to explore this space and propose novel adaptive alternatives. Be that as it may, our system does not deal with interspecies competition so we discarded this aspect of mutations.

Several implementations of the mutation process have been tested. All of them only deal with the production actions other than the energy resource production, assuming that all cells keep the capacity to gather and release all types of resources (minus eR).

As always, the simplest form is the random mutation. It affects the combination of resources required in the production (reactants) and the produced resource (product). These random mutations can be either beneficial or detrimental to the survival capacity of the new cell. Without any mechanism to stabilize the "genome" of the population in our simple setup, it is unlikely to converge towards any kind of coherent system. Emergence of cell-cell communication is improbable. Nevertheless, as seen in section 5.1.1, random mutations are able to induce the division of labor behavior in the system.

Communication at its simplest expression for our system would be for a cell to use of a non-essential resource as a request message to neighboring cells for an essential resource. For this to happen, cells need to be able to associate the presence of a resource in the environment (the request) with the release action of the essential resource (the response). Whatever the underlying mechanism used to implement the association process it does require some kind of correlations between resources present in the environment and the cell internal state. Mutations favoring strong correlations and removing uncorrelated actions could favor the convergence towards a communicating system. In biology, modern cells might have some mechanism to observe these resources correlations, adapt their behavior accordingly and pass that knowledge to the next generation. In early cells however, it is more probable that natural selection was the only drive to favor interesting correlations by "removing" cells that did not react to stimuli.

From the AMAS standpoint, to perform cooperative mutations, it is necessary to know on what bases cooperation has to be performed. Correlations between resources, the only entities of the system that can be observed by a cell, are thus necessary to favor the cooperation between the cells.

The next experiments explore this kind of cooperative mutations based on correlations.

5.2 Production Action Selection with Mutations

5.2.1 Correlations

In the long run, we want to have some resources influence the production of other resources inside the cell. As mentioned earlier we disabled the inhibition/activation of production actions by resources that are neither reactants nor product to avoid any conflict with the cooperative decision-making. Then, it is somehow necessary to include in the decision process this kind of resource cross influence. So, relationships between resources need to be established at some point in the simulation. To keep the local aspect of all the processes, these relationships need to be observed by each cell independently.

To build associations between actions and resources, each cell monitors the variations of resources in its environment. If over a time window a resource in the environment decreases at the

same time as another one, its correlation counter is increased. If on the contrary, one increases and the other decreases, its correlation counter is decreased. Over time, some correlations will become very positive, others will become very negative and the rest can be considered random noise. In this way, bonds between resources can be forged and used at the mutation stage. Although simple, this algorithm is quite sensitive to noise and its complexity grows exponentially with the number of resources in the environment.

As it is, this method is quite inefficient for our purpose and can detect a maximum of 48% of true resources correlations in a test system. In order to improve this score, some adjustments are added:

Correlation is only updated when a resource is released by the cell. That is, once a resource IR_s is released as ER_s a timer is started. After a predefined number of cycles nC that forms a time window, the history of variations for all the resources in the environment of the cell is analyzed (Figure 5-2):

- For each time t after the release of ER_s and for up to nC cycles:
    - If absolute variation is below a detection threshold, the score at time t stays unchanged.
    - If ER_i increases, the score at time t is increased by one.
    - If ER_i decreases, the score at time t is decreased by one.

To take into account memory decay or to allow for some plasticity of the correlations, every cycle the absolute value of the score is decreased by a small value. This value is adjusted so that one correlation observation disappears after 100 cycles.

The correlation of ER_s with ER_i is evaluated as the maximum unsigned score on the time window of nC cycles.

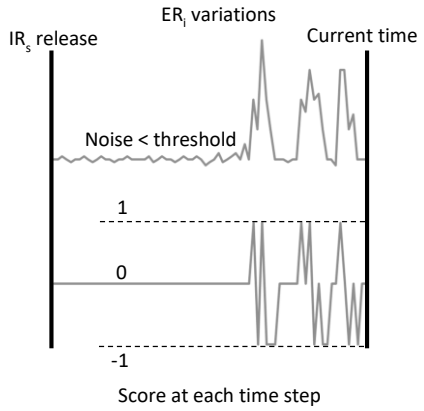


Figure 5-2 Correlation score between ER_s and ER_i.

This correlation calculation is more focused on the impact of the actions of a cell on its environment since only resources that are actually released by this cell are correlated. Decisions based on these scores are more likely to be pertinent to the local environment of the node. At the same time the drawback is that a lot of potential resource correlations are ignored, and novelty might be slow to appear with mutations using this algorithm.

Different strategies have been tested for the score adjustment when correlation between two resources are observed. For example, instead of using a +1/-1 score increase/decrease, the actual resources variations are used as a modulation. This seems like a better evaluation to adjust the score and get finely resolved correlations but this does not have a significant effect on the results presented in the following experiments.

5.2.2 Mutations Based on Resource Correlations

Based on the correlations described in the previous section, mutations are produced as follows:

- Each production action produces a single resource. The best correlation of this resource with other resources in the environment represents the score of this action. Then all production actions are sorted according to their score.
- The first one (with the highest correlated product) is kept since it has the most potential for the cell.
- Among the other production actions, the one with the lowest correlation is mutated.
- In order to explore potential interesting resources, the mutated action is changed in order to produce a resource for which no or few correlations have been established.

Once the system stabilizes, the production actions present in cells should all produce a resource correlated with a resource produced by its neighboring cells. In principle, these correlated actions should improve the cooperation between cells and might be the first phase of the emergence of communication strategies.

In order to test the usefulness of this mutation strategy, the following experiment was performed: On a 32×32 grid, 2 cell types coexist and 40 resources are available (20 in SetB). Cell type A uses resource R_{21} as an energy source and depends on resources R_2 and R_{39} to produce it. Five other production actions are generated randomly for each cell. These are the actions that will be mutated in subsequent cell generations.

Cell type A: Cannot release IR_{39} or IR_{21}

$$pAe_1: a_{11}.IR_2 + a_{12}.IR_{39} \rightarrow IR_{21}\ and\ a_{11} + a_{12} = 1$$

$$pAe_{i\in\{2\ldots6\}}: a_{i1}.IR_{i\,\in\,\mathrm{SetA}} + a_{i2}.IR_{j\,\in\,\mathrm{SetB}} \rightarrow IR_{k\,\in\,(\mathrm{SetB}-\{21,39\})}\ and\ a_{i1} + a_{i2} = 1$$

Cell type B does not mutate and its role is to provide type A cells with the required R_{39}. It is autonomous in terms of energy production (R_{21}) since it can produce it from two abundant resources (arbitrarily IR_0 and IR_{19}). It can also produce R_{39} from two abundant resources (arbitrarily IR_1 and IR_{18}), but only if the resource concentration $[R_{38}]$ on its node is above a threshold (0.01).

Cell type B:

$$pAe_1: a_{21}.IR_0 + a_{22}.IR_{19} \rightarrow IR_{20}\ and\ a_{21} + a_{22} = 1$$

$$pAe_2: b_{21}.IR_1 + b_{22}.IR_{18} \xrightarrow{[ER_{38}]>0.01} IR_{39}\ and\ b_{21} + b_{22} = 1$$

This setup makes A-cells dependent on B-cells. In order to survive, A-cells must release R_{38} in the environment. In response, B-cells have to release R_{39} that can be used by A-cells to produce R_{21}, their energy source. Using correlations, A-cells associate release of R_{38} with the subsequent release of R_{39}. Then the mutation algorithm described earlier favors the apparition and then the safeguarding of the production action associated with R_{38}.

At the start of the simulation, cells of the two types are placed randomly on the grid. Statistically, each A-cell has a good chance to be close to a B-cell. Some of the A-cells have a R_{38} production action from the start. These cells can quickly correlate R_{38} with R_{39} and survive long enough to transfer this good "gene" to their offspring. Once the dynamic equilibrium is reached, R_{38} production should be found in most if not all A-cells.

To be able to measure the efficiency of the mutation algorithm, the nursery is deactivated in this first experiment since it would interfere with the process. So, when a cell dies all its internal resources stay the same but for the energy resource which is set to allow it to survive for 200 cycles

(a necessary time to make some correlations). Also, the correlations are transmitted to the new cell as a mix of 20% of the correlations established during its previous life and 80% of correlations from all previous lives. This memory is essential for long term stabilization of the system. Finally, the mutation rate is 100% *i.e.* every new cell is mutated.

The results of this experiment are presented in Figure 5-3: B-cells are represented in green. A-cells that cannot produce R_{38} are in orange and in white when they can.

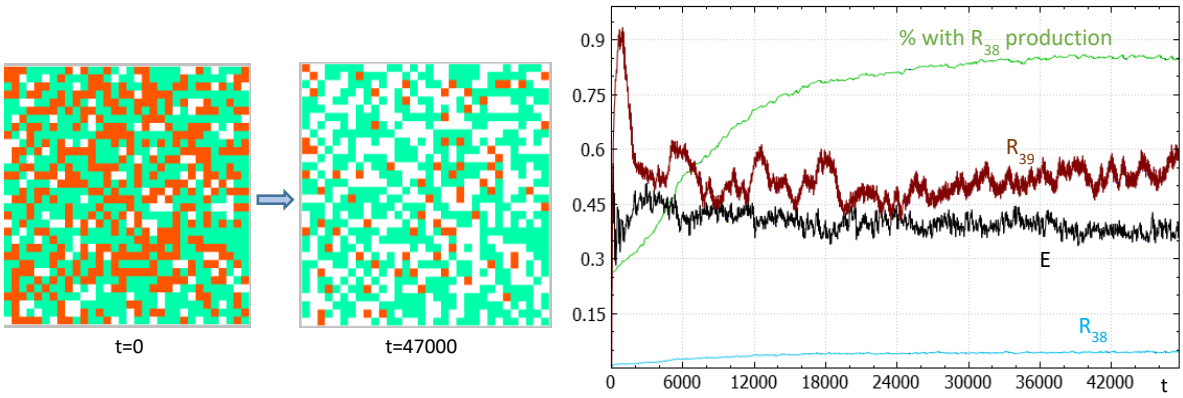


Figure 5-3 Acquisition of the R_{38} production action by correlation-based mutations

The system is able to reach a dynamic equilibrium where availability of R_{39} is sustained and let A-cells produce a continuous flow of R_{21} to survive. The action responsible for the production of the R_{38} signal is present in less than 30% of A-cells at the beginning of the simulation but raises steadily up to 85% at the plateau of the curve. The average A-cell age is around 10000 cycles which is indicator of a healthy system.

The same experiment performed using a random mutation algorithm is illustrated in Figure 5-4. On the same time scale as the previous experiment, the mutation process does not increase the population of cells producing R_{38}. This is to be expected since statistically there are as many cells that acquire the R_{38} production than there are that remove it every cycle. On the long term we could expect to observe an increase of this population since it gives these cells an increased fitness in this environment. But on the duration of this simulation, the effect is not observable. This indeed confirms that the cooperative mutation process using correlations accelerate the exploration of the system parameter space. The plateau value of 0.75 for $[R_{21}]$ is explained by the low lifespan of the cells (~500 cycles) and that each time a cell dies its energy resource R_{21} stock is renewed.

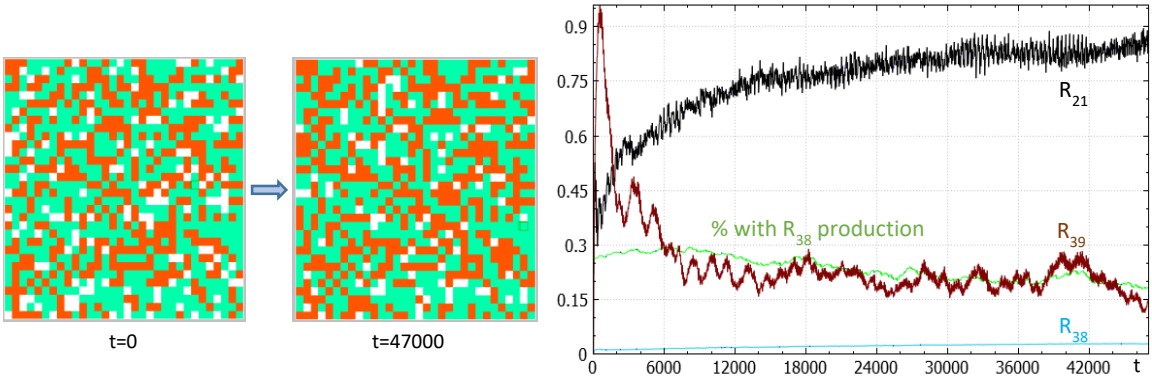


Figure 5-4 Acquisition of the R_{38} production action by random mutations

Although the mutation process looks promising by increasing the number of R_{38} producing cells threefold, it appears to reach a limit. Indeed, no parameter adjustment in the correlation algorithm

or the system dynamics can significantly improve the population of signal-producing cells above 85%. This is puzzling enough to investigate further this phenomenon.

5.2.2.1 *Modeling of the Mutation Process*

The evolution of the system can actually be modelled as a time series. By design of the mutation process, the percentage of cells that keep the R_{38} production action ("gene^{38}") is equal to the maximum correlation observed between R_{38} and other resources in the environment (hopefully R_{39}). This percentage is named Keepers^{38}. Getters^{38} is the percentage of cells that receive the R_{38} production action as a mutation.

At time t the number of dead cells that had the gene^{38} is $Dead_t^{38}$ and $Dead_t^{other}$ for the ones without the gene. These values are linked to the probability of a cell with or without the gene^{38} to die: P^{38} and P^{other}. These values are considered as constant in the model but are function of the survival advantage that the gene^{38} gives a cell. From these definitions we can express:

(1) $Dead_t^{38} = N_{t-1}^{38} * random(P^{38})$

(2) $Dead_t^{other} = \left(N^{total} - N_{t-1}^{38}\right) * random\left(P^{other}\right)$

Where N^{total} is the total number of cells in the system and N^{38}_{t-1} is the number of cells with gene^{38} at the previous time step.

From (1) and (2), the number of cells with gene^{38} at time step t can be expressed as:

(3) $N_t^{38} = N_{t-1}^{38} - \left((100 - Keepers^{38}) * Dead_t^{38}\right) + \left(Getters^{38} * Dead_t^{other}\right)$

That is, the previous number of cells with gene^{38} less the number of dead cells with gene^{38} that did not keep the gene during the mutation process plus the number of cells that acquired gene^{38} this cycle. A typical calculation of this time series is represented in Figure 5-5. After a few hundred cycles, the proportion of cells producing R_{38} stabilizes around a limit value. The curve matches quite well the actual one observed in Figure 5-3 for the full simulation system indicating that the model is representative of the phenomenon observed in the full simulation.

This series can easily be calculated for various values of the parameters P^{38}, P^{other}, Keepers^{38} and Getters^{38}. The limit values obtained for this calculation can tell us the relative influence of each parameter and provide some insights on the results obtained during the experiment with the complete system.

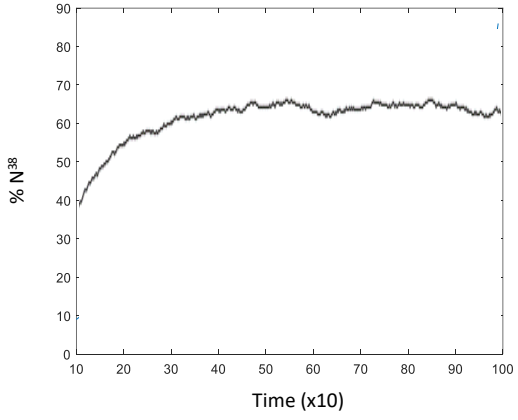


Figure 5-5 Example of evolution of N^{38} with time

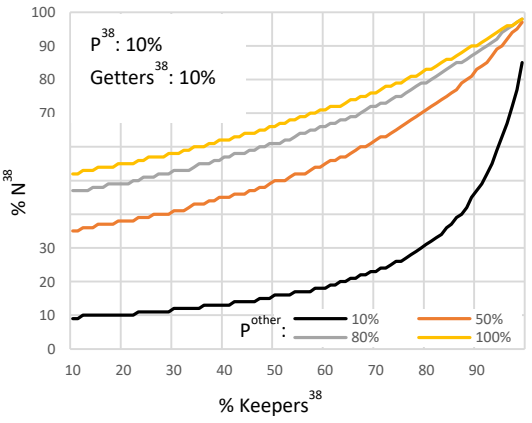


Figure 5-6 Evolution of the number of cells producing R_{38} with its maximum correlation

Figure 5-6 represents the proportion of cells producing R_{38} (N^{38}) as a function of the maximum correlation observed between this resource and other resources in the system ($Keepers^{38}$), at various death rates ratios (from 10% to 100%). Each value point is the average obtained for 1000 runs of equation (3) for 600 time steps (to reach the plateau observed in Figure 5-5). Very interestingly, the curves have an exponential shape and not a linear one apart at very high death rates for non-producing cells. This shows that to obtain a high percentage of cells with $gene^{38}$ the correlation must be very high and reliable. It also shows that the correlation is a poor lever to improve the global progression of $gene_{38}$ in the population since apart at very high correlation values, an increase in correlation has only a small effect on N^{38}.

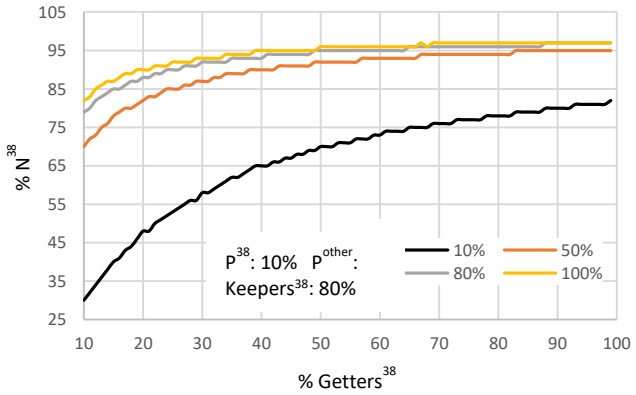


Figure 5-7 Evolution of the probability of getting $gene^{38}$ on N^{38}

Figure 5-7 shows the evolution of N^{38} as a function of the probability of getting $gene^{38}$, $Getters^{38}$, during the mutation process. This parameter has a more interesting influence on the population evolution than the correlations since it has a near linear shape between 10 and 30% to reach a plateau around 60%. In the mutation algorithm, this parameter is difficult to alter since it is related to the lack of information a cell has about R_{38} correlations with other resources. This is very dependent on the history of the system in general and on the local node in particular.

The last parameter to investigate is the death probability for cells that do not produce R_{38} (P^{other}). Figure 5-8 shows that the profile is very similar to the influence of $Getters^{38}$. The difference is that it is much easier to change this parameter than the previous one. Indeed, to increase P^{other}, $gene^{38}$ must be essential to the survival of the cell. This is possible to achieve in the simulation, for example by increasing the rate of ER_{39} removal from the system. This way, ER_{39} will not be able to diffuse far from its emission point and will benefit more to cells requesting it by releasing IR_{38} and less to cells unable to produce IR_{38}.

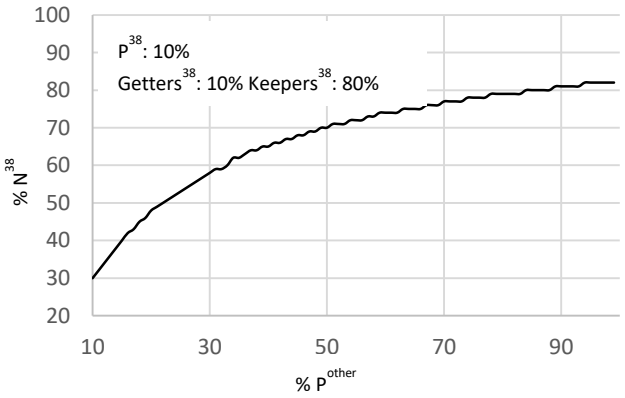


Figure 5-8 Influence of P^{other} on N^{38}

Finally Figure 5-9 presents a synthetic view of the parameter space of the N^{38} evolution model. Light shade zones represent population with dominance of the gene^{38}. These zones are accessed only with high values of Keepers^{38} or Getters^{38}. Also, there occupy more space when P^{other} is high.

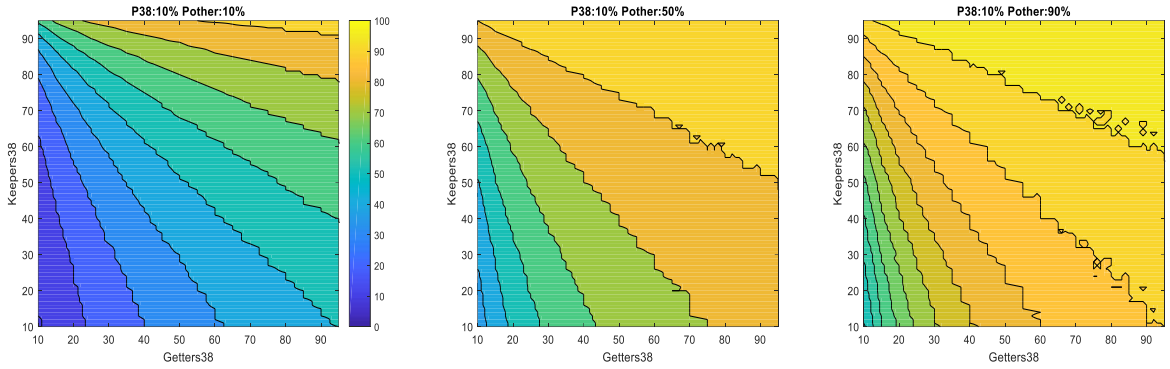


Figure 5-9 Proportion of N^{38} (color scale) as a function of %Getters^{38}, %Keepers^{38} and P^{other}

The main conclusion of the study of this model is that the correlation accuracy, which is linked to gene^{38} upholding, is not an easy target for improving the final outcome of the system. Only very high accuracy can ensure the retention of the gene^{38} and this is very difficult to achieve in a noisy environment. Does this conclusion have any relevance for real life systems? This is far from clear since as far as we know, mutation in real cells is not dependent on resource correlations. Nevertheless, it is interesting to note that signal/response correlations in cells are not always perfect but this does not reduce the capacity of the cell to keep the corresponding genes/function. This might be interpreted as the fact that after a certain accuracy value, it is not worth (energetically) to invest in better systems but it is preferable to alter other parameters like P^{other} in order to keep a useful function.

5.3 Conclusion on Correlation-based Mutations

Although mutations induce novelty in the simulation, they also bring problems. In particular, at the global level it becomes difficult to foresee the direction of the evolution of the system and thus to control it. Even in a simple system as CoCell, it is rather complicated to have the cells behave as we would like them to and more often than not, mutations give them the ability to find an original solution to survival that was not predicted or desired. Like in real life, mutations provide cells with a way to escape constraints of the environment or turn them into strength.

Other mutation algorithms would be possible to implement and might yield better results on different scenarios but our investigation is focused on signal/response thus resource correlations are meant to play a role at some point. Since the performance of this correlation-based mutation process is acceptable, it is used in the next stage of the simulation CoCell3 where the emergence of communication is investigated in more details.

It is noteworthy to mention that these correlation-based mutations do not perform better than a random mutation process in a situation where resource correlations do not have any impact on cell fitness or if all resources are independent. For example, in section 4.3.2, the system using 10 different cell types and 20 resources can be simulated with the correlation-based mutation process. It appears that since there is no correlation in the system, the equilibration is neither reached sooner (*i.e.* never) nor more stable than when using random mutations.

CHAPTER 6. COCELL3: TOWARDS EMERGENCE OF COMMUNICATION

Dans ce chapitre, la communication intercellulaire, la troisième étape de la simulation est étudiée. Toutes les principales méthodologies ont été introduites dans les itérations précédentes, CoCell1 et CoCell2, comme la prise de décision coopérative, le remplacement des cellules mortes et les mutations. Nous examinons ici les conditions nécessaires pour observer l'émergence de la communication dans un système où l'interdépendance entre les types de cellules est forte.

Certaines hypothèses importantes doivent être exprimées et mises en œuvre qui sont les conditions préalables pour qu'une communication se produise entre les cellules.

Premièrement, comme nous l'avons déjà mentionné, la libération d'un signal par une cellule est un processus qui a un coût pour la cellule en termes de travail et d'énergie. C'est-à-dire que le signal doit être synthétisé, stocké et libéré dans l'environnement. Pour qu'un tel processus soit maintenu au niveau cellulaire, il doit avoir une rétroaction positive sur la durée de vie de la cellule et ce gain doit être supérieur à tout autre type de mécanisme qui pourrait jouer le même rôle et améliorer la survie de la cellule. Une forte interdépendance entre les types de cellules est un moyen naturel, observé dans la nature, qui favorise un tel comportement. Mais est-ce suffisant ?

Deuxièmement, la libération d'un signal doit être liée d'une manière ou d'une autre à un besoin de la cellule qu'elle ne peut satisfaire par elle-même. Il doit donc exister un mécanisme qui crée et renforce ce lien entre le besoin et le signal. Pour émerger à partir d'actions apparemment aléatoires de collecte, de distribution et de production, le lien peut être établi par des observations de soi et de l'environnement. Les influences croisées sont alors établies et renforcées si elles réussissent à améliorer la survie des cellules et du système dans son ensemble. En termes de cellules réelles et d'évolution, ces corrélations sont probablement créées par le bruit aléatoire pur et renforcées par la sélection naturelle. Comme nous l'avons déjà mentionné à plusieurs reprises, dans le contexte d'une simulation, le hasard est un mauvais allié puisque nous n'avons pas le temps ni la puissance de calcul pour l'utiliser à son plein potentiel. Nous devons donc utiliser d'autres moyens de produire les mêmes résultats de manière plus efficace et sans introduire de biais. Ici, nous avons décidé d'évaluer les corrélations temporelles entre les ressources afin d'identifier les liens potentiels entre ces ressources.

Enfin, la parenté entre cellules et la coopération utilisée comme axiome, permettent de faire la dernière hypothèse : toutes les cellules se comportent de la même manière. Il s'agit d'une prémisse solide qui a d'importantes implications simplificatrices. En effet, construire le comportement coopératif d'une cellule en supposant que les cellules voisines travaillent de la même manière est un grand avantage pour l'élaboration d'une stratégie de communication.

In this chapter, intercellular communication, the third step of the simulation is studied. All main methodologies have been introduced in the previous iterations, CoCell1 and CoCell2, like cooperative decision-making, nursery and mutations. Here we look at the conditions necessary to observe the emergence of communication in a system where interdependency between cell types is strong.

Now, some important assumptions have to be expressed and implemented that are prerequisite for actual communication to occur between cells.

First, as mentioned earlier, the release of a signal by a cell is a process that has a cost for the cell in terms of work and energy. That is, the signal needs to be synthesized, stored and released in the environment. To allow such a process to be maintained at the cellular level, it must have a positive feedback on the life-span of the cell and this gain must be superior to any other kind of mechanism that could play the same role and would improve the cell survival. Strong interdependence between cell types is a natural way, observed in Nature, that promotes such behavior. But is it sufficient?

Secondly, the release of a signal must be somehow linked to a cell need that it cannot satisfy on its own. So there must exist a mechanism that creates and then strengthens this link between need and signal. To emerge from seemingly random gather, release and production actions, the link can be established by self and environment observations. Cross influences are then established and reinforced if they prove successful to improve the cell survival. In terms of real cells and evolution, these correlations are probably created by pure random noise and reinforced by natural selection. As mentioned several times already, in the context of a simulation, randomness is a poor ally since we do not have the time and computing power to use it to its full potential. So, we must use alternative ways to produce the same results more efficiently and without introducing any bias. Here, we decided to evaluate time correlations between resources in order to identify potential links between resources.

Finally, kinship between cells and cooperation used as an axiom, enable to make the last assumption: all cells behave in the same way. This is a strong premise that has important simplifying implications. Indeed, building the cooperative behavior of a cell assuming that neighboring cells work in the same way is a great advantage for the elaboration of a communication strategy.

6.1 Cell Decision-making Process

The decision-making process used in CoCell1 was purely based on criticalities of resources inside the cell and on the node. A cell tries to help the most critical resource using the possible actions at its disposal. The mutation process used in CoCell2 only modified the set of actions present in the cell and was unable to modify the way a cell took its decisions so the decision mechanism was not changed and thus was not able to adapt to new circumstances.

In order to use the correlations observed by the cell, this decision making process requires to be updated. In this last stage of the study, we try to make it more compatible with the use of inter cellular communication but here also these changes need to be carefully weighed against the bias they introduce about communication.

The approach chosen for designing this last simulation, CoCell3, is to perform three iterations. In the first one, we chose to design a process that actually works and leads to the expected emergence of communication regardless of the bias introduced to this aim. Then, in the second iteration we would try to remove most of this bias. Finally, the decision process could become itself a target of the cooperative mutation phase and be changed at the same time as the cell actions.

Ideally, in this last iteration, we would like the communication bias to disappear and be replaced by a fully cooperative and unbiased behavior.

The decision process of a cell uses the correlation information it has gathered in order to request resources it lacks for surviving.

In a cell colony where cells are interdependent, these correlations are as important for a cell survival as the actions this cell can perform. So this kind of information is transmitted during cell division. In real cells, this can be done through some pathway regulation proteins encoded in DNA or in the delicate balance of proteins and metabolites present in the cytoplasm of the cell. In the simulation, correlations observed during the cell life are added to the correlations observed by its ancestor cells when they are passed on a new generation. Furthermore, during its life, a cell bases its decisions on a combination between correlations it observed and correlations passed on by its mother:

$$Actual\_Correlations = Factor * Mother\_Correlations + (1 - Factor) * Observed\_Correlations$$

In CoCell3, the decision process performed by a cell is divided into two modules: the first one expresses the egoistic part of a cell, and the second one, represents its altruistic part.

The first module tries to improve the stocks of important resources to ensure long-term survival of a cell and will release "useless" resources as "signals" whenever it cannot satisfy a need by using gather or production actions. The selection of the resource to release is based on the correlations this cell observed between this resource and the needed one. This is the first part of the bias introduced to favor communication since we favor the emission of a help signal.

The second module used by a cell detects the needs of surrounding cells assuming they work in the same way as itself, by sending a request when in need of something. This is where the second part of the bias towards communication is introduced since we implement a response mechanism. So, whenever a cell observes a resource present in the environment that is correlated with one of the resources it can produce, it is assumed that this represents a request generated by one of its neighbors for that resource and consequently releases it.

The action decision proceeds as follows:

- Internal and external resources are ranked according to their criticality as in previous versions of CoCell.
- Two actions are performed by the cell, one selfish and one altruist:
    - Egoistic: From the most critical internal resource to the first non critical one.
        - Help the critical resource if an action to do so is available: For example, gather or produce the resource.
        - If no direct action is possible for the critical resource, perform an action to help its most correlated resource: This can either be to release in the environment or to produce the most correlated resource, as a surrogate for the critical resource, if the surrogate external criticality is high enough (so that it will appear as a concentration peak in the environment). This represent a signaling mechanism since the cell uses a surrogate to request help from the environment.
    - Altruistic: From the least critical external resource to the first critical one.
        - If a resource is non critical in the environment and is correlated with a critical resource the cell is able to produce, release this resource if it is available. This is a response in a signal mechanism.

Following these rules, a cell will acquire important resources and use surrogate resources as signal/response when it cannot directly impact a critical resource. Furthermore, a cell does not release resources spontaneously unless it detects what appears to be a request for it.

Using this modified decision process, the conditions are more favorable to observe the emergence of communication between interdependent cells but obviously it includes a bias towards this end.

6.2 Modification of Cell Behavior

Several improvements (or at least thought of as improvements) are added in CoCell3 when compared to earlier versions. Each time a change is applied, its influence is evaluated on the system behavior in the scenario described in earlier chapters (reaching equilibrium for several cell types or evolving a population where all cells acquire a critical action). If a modification renders the system unable to survive one of these scenarios or deeply alters its behavior, this modification is rejected. Among the changes that are kept, the most notable ones are detailed below.

6.2.1 Correlations

The correlation algorithm is altered in a way that distinguishes signals from responses to signals.

For a given resource R_i, two sets of correlations are calculated over a time window. The first one refers to the potential of the R_i to act as a signal to request resource R_j. The second one refers to the potential of R_i to be an adequate response upon reception of signal R_j. Unlike the correlations described in section 5.2.1 which were evaluated after the release of R_i, these new correlations are purely based on the observation of the resources variations in the environment of the cell regardless of the cell actions.

These two evaluations are performed as follows:

- Every cycle, the correlations fade by a specific small factor.
- Correlation of resource R_i as potential signal for resource R_j is evaluated in this way:
    - For each peak of R_i observed in the correlation time window (time t_1, see Figure 6-1), peaks of R_j are searched after t_1.
        - If R_j increases at $t_2 > t_1$, R_i is a potential signal for R_j with a latency of $\Delta t = t_2 - t_1$:
            - $Signal\_Correlation_{ij}^{\Delta t}$ is increased by a given value.
        - If R_j decreases:
            - $Signal\_Correlation_{ij}^{\Delta t}$ is decreased by a given value (which can be different from the increase value).
    - The final signal correlation between R_i and R_j is given by:

$$Signal\_Correlation_{ij} = \max_{\Delta t}\left(Signal\_Correlation_{ij}^{\Delta t}\right)$$

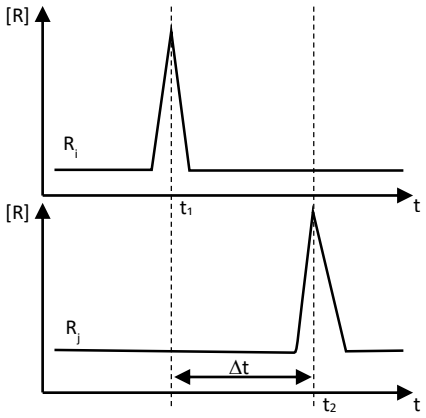


Figure 6-1: Signal correlation between R_i and R_j.

- Correlation of resource R_i as potential response for resource R_j:
  - For each peak of R_i observed in the correlation time window (time t_1, see Figure 6-2), peaks of R_j are searched before t_1.
    - If R_j increases at $t_2 < t_1$, R_i is a potential response for signal R_j with a latency of $\Delta t = t_1 - t_2$:
      - $Response\_Correlation_{ij}^{\Delta t}$ is increased by a given value.
    - If R_j decreases:
      - $Response\_Correlation_{ij}^{\Delta t}$ is decreased by a certain value.
  - The final Response correlation between R_i and R_j is given by:

$$Response\_Correlation_{ij} = \max_{\Delta t}\left(Response\_Correlation_{ij}^{\Delta t}\right)$$

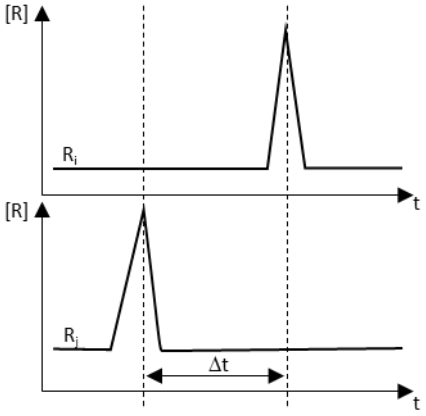


Figure 6-2 Response correlation between R_i and R_j.

Although this correlation calculation is more sensitive to noise than correlations based on resource release, it is based on more observations since there are more events in the environment than release actions events.

Also, if proven efficient, this method can be further improved since if signal A is found to be highly correlated with response B, then B must be found to be highly correlated as a response to A. Thus, the detection of signal/response could be enhanced by comparing these two values to filter out ambient noise.

6.2.2 Weighted Resource Amount

In CoCell1 and CoCell2, the amount of resource processed (R^{Δ}) by each type of action is only limited by the amount of resource available and a global maximum (max^{Δ}) (see section 3.2.5.2). That is, actions might in some case be excessive in their effects since the amount of resource modified can be greater than the quantity needed to "help" the resource selected by a cell.

Although this is not a major problem when the conditions in the environment are not too demanding, it can become challenging in some more difficult circumstances since the extra work consumes energy.

Also, it is not very close to the way modern cells work. Indeed, very early versions of cells were probably just a collection of chemical reactions occurring in the defined space of the cell, and would process chemical reactants until there were none left or balanced by the rules of chemistry. But, in modern cells, regulated catalysis and feedback loops finely tune the output of each and every chemical reactions taking place inside the cell and it is always (or most of the time?) the exact amount required for the optimal functioning of the cell.

Since CoCell3 deals with more advanced cellular models that are engaged in the transition from monocellular to multicellular organisms, the actions ought to be more parsimonious. In order to achieve more regulated actions, criticalities of the resources are used in a form of mass action law mechanism:

- For a release action, the amount of resource Ri to release is evaluated as before but is then adjusted by a factor $\mathrm{f} = \frac{\mathrm{ECrit_i}}{\mathrm{ICrit_i}}$ only if f < 1. So if a cell decided to release a resource but this latter is more critical inside the cell than in the environment, the final amount released is decreased.
- For a gather action, the gathered amount is weighed by a factor $\mathrm{f} = \frac{\mathrm{ICrit_i}}{\mathrm{ECrit_i}}$ only if f < 1. Thus, a cell will not gather too much from the environment and deplete it when it already has some of the gathered resource in stock.
- For a production action that only uses internal resources, the weighting factor is $\mathrm{f} = \frac{\#reactants * \mathrm{ICrit_{product}}}{\sum_{\#reactants} \mathrm{ICrit_j}}$ only if f < 1. Thus, if the reactants of the production action are in low supply, the amount produced will not deplete them too much.

This new mechanism does not modify profoundly the overall behavior of the system in any of the experiments performed. In situations where the system is able to reach a dynamic equilibrium using the old amount evaluation, it is still able to do so, and when it was not possible to reach a stable state it is still not able to reach it. This is probably due to the simplicity of our model that is not sensitive enough to be altered by this modification. It is probable that a finer simulation of the chemistry would be needed to observe the influence of finer production amounts. Nevertheless, this is a somehow more satisfactory mechanism closer to what could happen in a real cell.

It is still possible to improve this factor evaluation further. For example, by using extrapolated resource criticalities after the reaction has taken place and adjusting the factor accordingly. This would actually be equivalent to solving a simple differential equation to find the optimal factor. Although more precise, it is not clear if that would have a critical impact on the system behavior. But it would certainly increase the computation time per simulation cycle.

6.2.3 Diffusion and Cell Actions

Another structural modification in CoCell3 concerns the gather and release action behaviors. As mentioned earlier (see section 3.2.5.2), gather and release actions act on resources present on the node of the cell performing the action and also on the neighboring nodes. Thus, a release action distributes a resource on nine nodes (in the standard setup used in the presented experiments, but this is a parameter). Similarly, a gather action harvests resources from the same nine nodes. As mentioned in section 4.1.3, since several cells can decide to gather the same resource on the same shared neighboring nodes, a concurrent resource access approach needs to be used. This increases the difficulty of the algorithm, requires a multistep approach and cannot be easily parallelized.

A more fundamental problem with these release and gather action behaviors lies in the hidden resource diffusion that it involves. Transport of resources across the system could be performed using only gather and release actions at a speed that is indifferent to the diffusion parameters of the resources.

To avoid the interference of the cell actions with the passive diffusion mechanism, their operating mode is modified in CoCell3. A gather or release action only acts on the node resource content of the cell performing the action. In this way the algorithmic complexity decreases, parallelization is easier and the diffusion is perfectly decoupled from cell actions.

6.2.4 Mutation Process

Since the cell decision process includes an egoistic and an altruistic part, the mutation process is modified in order to enhance both these aspects in the daughter cell. Thus, the new mutation algorithm includes two parts, one for the egoistic part of the cell and a second for the altruistic part.

The process is as follows:

- For the selfish part:
    - The action that produced the resource with the lowest signal correlation to any other resource is selected for replacement: pA_{low}
    - The resource that is not already produced by the cell and has the highest signal correlation score is selected: R_{high}
    - A production action that generates R_{high} replaces pA_{low}.
- For the altruistic part:
    - The action that produced the resource with the lowest response correlation to any other resource is selected for replacement: pA'_{low}
    - The resource that is not already produced by the cell and has the highest response correlation score is selected: R'_{high}
    - A production action that generates R'_{high} replaces pA'_{low}.

6.3 Experiments

A system similar to the one described in 5.1.1 is first used to test the new decision process. In this system, a single cell type can produce energy from two rare resources from SetB, R_a and R_b. When a cell produces both of these rare resources, an energy penalty is applied. During the simulation, energy constraints and "natural" selection favor the apparition of two different cell subtypes, each producing only one of the required rare resources. In this context, the biased

decision process is expected to orient the system in a way where R_a and R_b are only released in the environment when requested by cells needing them for their immediate survival. By using the described resource correlation mechanism, the release of R_a and R_b is conditioned by the presence of another resource that takes the role of a signal.

As shown on Figure 6-3, it appears that this scenario cannot easily be used to observe the emergence of signal/response behaviors. Indeed, the correlations tend to focus on relations between R_a and R_b only. So, to obtain R_a, a cell releases R_b and reciprocally a cell that requires R_b releases R_a. This behavior is actually the most efficient one possible in this particular setup. Using a different resource as signal to request either R_a or R_b just adds extra energy consuming steps to produce and release the signal. By using R_a as a request signal for R_b and reciprocally, the cells are able to kill two birds with one stone. They actually inform their neighbors that they are in need of help and at the same time they help cells that requested R_a.

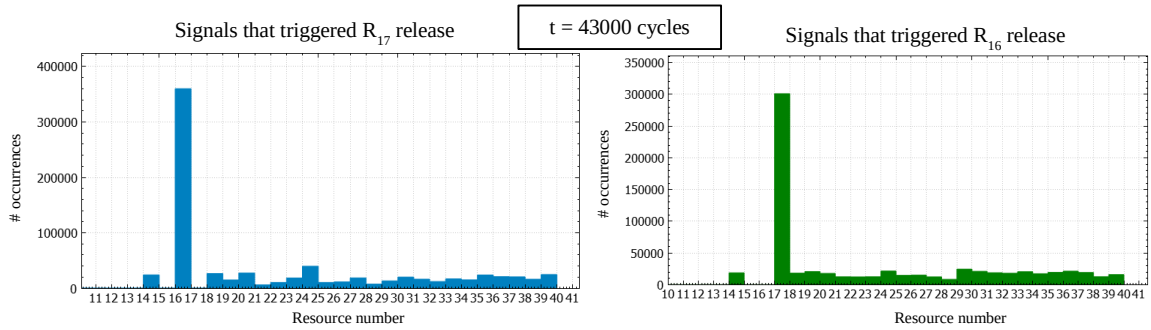


Figure 6-3 Distribution of resources used as signals to request essential resources R_{16} and R_{17} over a 43000 cycles simulation.

So, in order to observe non trivial communication, a slightly more complex system is designed. Instead of a single ancestor cell, there are three of them. Each one depends on the same energy resource R_e, but they use different combinations of two rare resources (among R_a, R_b and $R_c \in$ SetB) to produce it. The three energy production actions are designed as follows:

$$Cell\ type_1\ epA_1: a_{11}.IR_a + a_{12}.IR_b \rightarrow IR_e\ and\ a_{11} + a_{12} = 1$$
$$Cell\ type_2\ epA_2: a_{21}.IR_b + a_{22}.IR_c \rightarrow IR_e\ and\ a_{21} + a_{22} = 1$$
$$Cell\ type_3\ epA_3: a_{31}.IR_c + a_{32}.IR_a \rightarrow IR_e\ and\ a_{31} + a_{32} = 1$$

Each cell type can produce one of the essential resource for the two other cell types. So, each one depends on the two other cell types to survive. This setup is much more demanding of the cooperative algorithms to reach a dynamic equilibrium but should avoid trivial communication like bartering of essential resources.

Each cell instance possesses nine other random rare resource production actions. The energy production cost of a resource is inversely proportional to its number. This is to introduce some differences between resources and try to favor the use of low cost resources as signals.

Instead of inducing cell type interdependence using an energy cost penalty as done in previous experiments, here the productions of R_a, R_b or R_c are disabled if the resource is used by the cell to produce R_e. This is a bias implemented to accelerate the convergence of the system. To be sure to avoid the simple essential resource exchange observed in the two cell types experiment mentioned earlier, the same R_a, R_b or R_c cannot be used as signals by the cells. And to avoid simple energy resource sharing, R_e cannot be released in the environment. This could be a true situation if the compound is completely unstable in the environment conditions.

Instances of each cell type may or may not produce the vital resource required by the other two cell types. If not, the cell instance becomes a cheater and endangers the stability of the system. The node cooperative nursery and the cell cooperative mutations should accelerate the process of cheater disappearance in favor of cooperative cells.

Every cycle, each cell observes its environment and records the resources variations and correlations over a time period in the immediate past, as described in section 6.2.1.

The conditions of this simulation are as follows: The system space is 32×32, the number of resources is 40 (12 in SetA and 28 in SetB), the mean cellular life-span is set to 2500 cycles, the correlation mixing factor is 0.8. The energy resource R_e is IR_{15}, and its possible constituent R_a, R_b and R_c are respectively R_{16}, R_{17} and R_{18}. The initial proportion of cooperative cells (that can produce the essential resource for the two other cell types) is set to 0.40. The three cell types have initially an equivalent number of instances that are placed randomly in the system space. The time window length to evaluate resource correlations is set to 15 cycles. The correlation fading is set so that a single correlation observation is forgotten after 100 cycles.

The experiment is run for around 350000 cycles. The results presented below represent a single representative experiment. Using different random seeds, the evolution of the system, although locally different, is globally the same, but the resources used by the system as signals are different each time.

Figure 6-4 presents the evolution with time of resources from SetB. A dynamic equilibrium is reached early and maintained, although some periods in the history of the system appear more chaotic than others. It is interesting to note that several resources eventually fade out of the system (R_{19}, R_{38} and R_{23}) as cells "find" them unessential and not correlated to important resources (Figure 6-4 a).

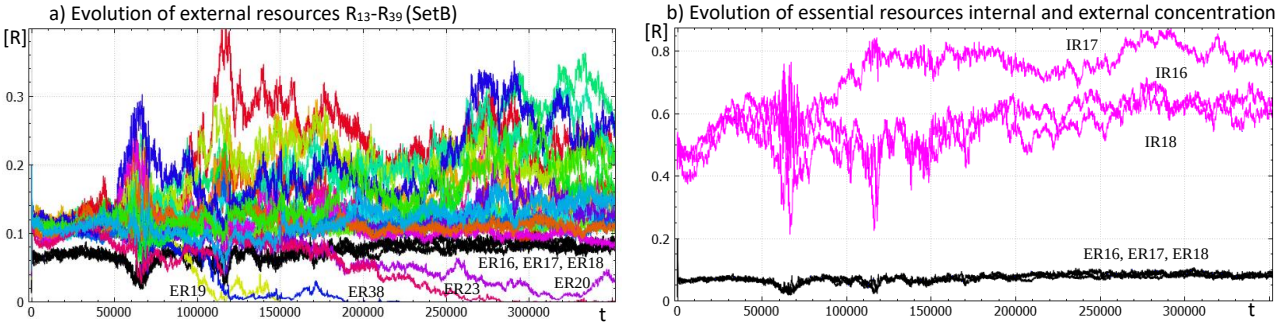


Figure 6-4 Evolution with time of mean concentration for rare resources. Essential resources ER_{16}, ER_{17} and ER_{18} are colored in black.

When comparing the essential resource concentration inside and outside the cells, it appears that healthy stocks are maintained inside the cell. This demonstrates two things: Firstly, the exchange of resources between producing and non-producing cells works well. Secondly, the egoistic part of the decision-making algorithm favors the survival of the cell by promoting the storage of important resources, as intended.

Figure 6-5 presents the cell type distribution shortly after the start of the system and at the end of the simulation. They are very similar spatially and quantitatively as shown in Figure 6-5 a) and b). This shows that the system is fairly stable. Although at some points of its history the situation was quite different, the cooperative processes were able to rebalance/save the system.

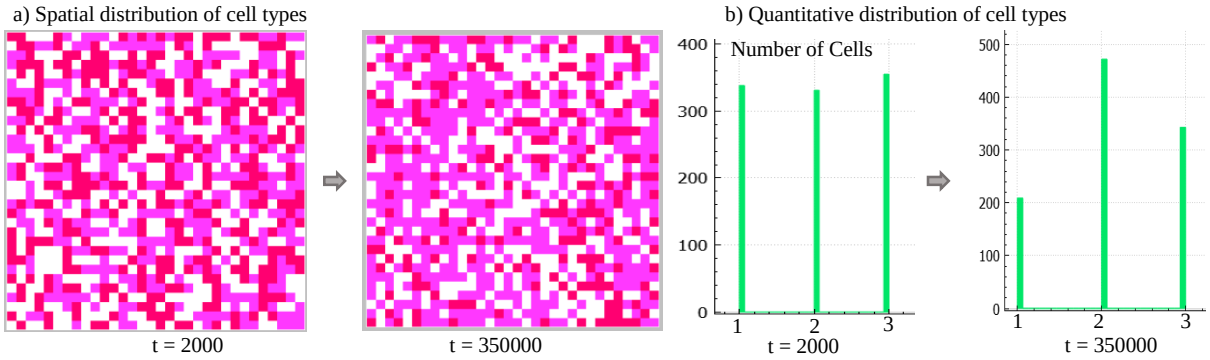


Figure 6-5 Cell type distribution in the system. Types 1, 2 and 3 are respectively colored in red, purple and white in the spatial distribution.

As in most of the simulations presented so far, there is no visual pattern emerging from the simulation. This can be the result of the limited size of the system but it is probably more the consequence of random noise in the system and the strong requirement for each cell to be close to two complementary cells.

Apart from the establishment of a dynamic equilibrium, two important parameters of the system are the average cell age and the degree of cooperativity. Figure 6-6 shows the evolution of these two parameters with time. The degree of cooperativity is measured by the ratio between the number of cells that can produce the essential resources required by the two other cell types and the number of cells that are "cheaters". In Figure 6-6 a) the spatial distribution of cooperative cells shows that they are more or less evenly distributed through the system and that it does not change much with the age of the system although there is a better coverage at t=350000 since there are more cooperative cells present (53% *vs* 40% at the beginning). The evolution of the ratio is shown in Figure 6-6 b). At several occasions, the ratio went down with a minimum of only 10% cooperative cells present in the system around t=65000. Nevertheless, with time the ratio increases very slowly. It is not certain that it would eventually reach 100% as time passes. It might be that a 100%-cooperative population is not required to have enough cells provide others cells with the necessary essential resources. This is probably very dependent of the system conditions. If they are forgiving, few cooperative cells can possibly produce enough resources for a large system. The results we observe here tend to show that we have lax surviving conditions. This will be further highlighted later.

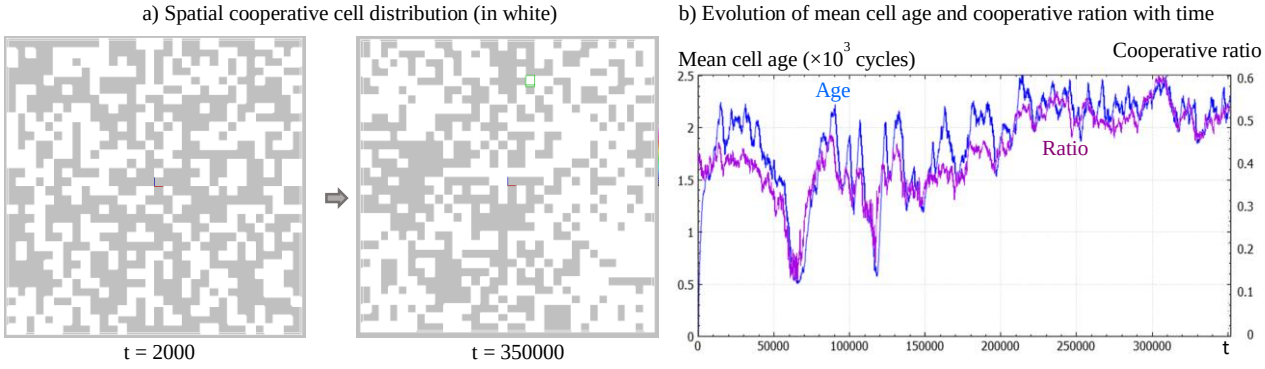


Figure 6-6 Evolution of mean cell age and cooperative ratio with time.

The second important parameter is the average cell age. It should be around 2500 cycles as set in the simulation parameters. Figure 6-6 b) displays its evolution with time. There are important fluctuations during the life of the system but at the end of the simulation the average cell age has

reached a value of 2100 cycles. This means that although most cells live up to their life-expectancy, some cells die prematurely from other causes which results in a lower than expected mean value. These other causes are mainly lack of energy, but also on some occasions uselessness, when a cell cannot perform any useful action for too many consecutive cycles and performs an apoptosis action.

An interesting aspect of the average age curve is that it is very highly correlated with the cooperative ratio. The two parameters appear to evolve together although they are slightly offset from each other. The cooperative ratio appears to evolve in one direction or another just before the average cellular age evolves in the same direction. If any conclusion can be drawn from this apparent correlation, it would be that directly or indirectly, the cooperative ratio influences the survival ability of cells. This seems coherent with the system design since more cooperative cells should correspond to more essential resources available to share, thus an increase of cell survival potential and age.

The most interesting information obtained in this experiment is shown in Figure 6-7. This figure shows the distribution of signals used by cells that resulted in the release of the specific essential resource on a period of 400 cycles around the indicated time. At time t=2000 cycles, correlations have been evaluated more than 100 times and the average cell age is already close 1500 cycles. Each cell has had the time to send requests for essential resources and to answer similar queries.

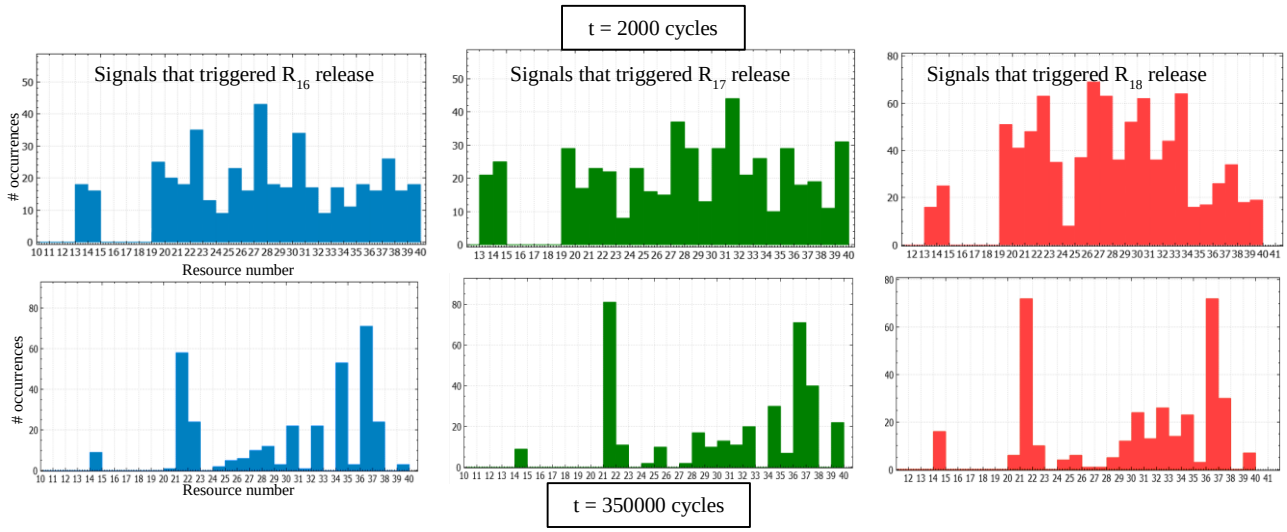


Figure 6-7 Distribution of resources used as signals to request essential resources R_{16}, R_{17} and R_{18} at the beginning and at the end of the simulation.

What is observed in Figure 6-7 upper row for t=2000 cycles, is that most rare resources are used to request for essential resources (apart from R_{15}, R_{16}, R_{17} and R_{18} which are forbidden). Some resources might appear to be used more often than others on this slice of 400 cycles, but in consecutive slices other resources would appear to be used more. On average, no resource appears to be even a weak consensus among the population of cells. After 350000 simulation cycles and several near death experiences for the system, the picture is quite different (Figure 6-7 lower row). Indeed, most of the rare resources are rarely or never used as signals. The distribution profile between the three essential resources is very similar with R_{21} and R_{36} as main contributors. There are some differences too since for R_{16}, R_{34} is also used often. At first sight this could be interpreted as emerging channels of communications between cells from a (nearly) undifferentiated set of resources. A somehow interesting result (although long to get) for this system.

It is noteworthy to observe that the main signaling resources are the same for the three essential resources. This can be explained in several ways.

When a given signal is used, it might happen that over the correlation window (15 cycles in this experiment), two different resources are released by different cells as responses. A cell observing the medium considers the two responses as equally valid. There is no hierarchy between long delay and short delay response in the current correlation algorithm. Consecutively, this cell would use the same signal to request two different resources. This mechanism can lead to the observed system where one or two signals are used to request any kind of resource. It would be equivalent to a simple and generic "help" signal. In an undemanding environment, this unspecific request can be enough for each cell to obtain what it requires to survive.

Another interpretation of this result is an increased cooperativity between cells: When a cell sends a signal for one of its essential resources, say R_{16}, it will elicit a response in the surrounding cells. Some may respond with R_{16}, some with R_{17} or R_{18} or some other non-essential resource. The result is that the solicitor cell gets some R_{16} but also gets its second essential resources at no further expanse. Furthermore, the third essential resource released can be useful to another cell in the surrounding area. So, in a setup where resources are quite cheap to produce, a generic request signal is efficient since it helps the cell sending the signal and also other cells in the neighborhood. This is a perfect cooperative action since the cell can get help as well as provide help to others in a single action. If the resource released in the medium in response to the request are not used by any other cell, this does not represent a big waste and does not endanger the viability of the system when resources are cheap to produce.

The fact that after 350000 cycles there are still at least two main signals for the three essential resources is explained by the fact that there are some heterogeneities in the production actions available to cells. Some cells are able to produce only R_{21} and others produce only R_{36}, and their offspring inherit this trait. At some point, it is possible that one of the resource will be favored and will become the only generic help request signal. If production energy cost has any influence in this lengthy process, R_{36} should become the unique signal since it is cheaper to produce than R_{21}. It is also possible that at some point each signal becomes specific of a single resource, initiating the development of a richer "vocabulary". It is interesting to note that in different runs of the same simulation, only one signal stands at the same system age, and in others three resources are still used.

These results give rise to a question about this system setup: How easy and good is it for a cell to live in it? To answer it, the calculated correlations are replaced by random numbers and the system is run using the same conditions as before. And the result is that the system is very lax indeed as presented in Figure 6-8. The system reaches a dynamic equilibrium, and the average cellular age of about 2200 cycles is close to the defined limit and also very similar to the mean age observed in the system using correlations. The main difference is observed in the signals used to trigger release of essential resources like R_{16} (Figure 6-8 b): All possible resources are nearly evenly used by the cells since correlations at each time step are random. The conclusion is that even without correlations to guide it, the system can survive with decisions based on random noise. In this context, communication only adds a very slight survival advantage. This would explain why, in the experiment using calculated correlations, communication emerges so slowly in the system.

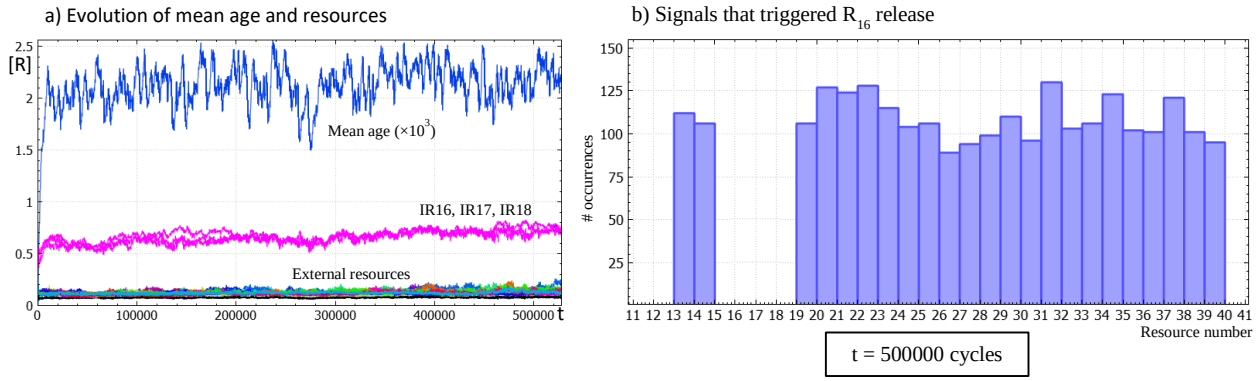


Figure 6-8 Same simulation as in Figure 6-4 but using random correlations.

This result is disappointing since it would be much more interesting to show that the emergent communication system is essential to the system survival and that switching it off destabilizes the system and eventually results in its death.

The problem is that it is very difficult to setup the system in a way that is both stringent enough to require communication to survive and easy enough to let the cells survive during the time an efficient communication protocol emerges between the cells. Maybe a multistep scenario is necessary where conditions are mild at the beginning and become increasingly harsh as time passes. In this case, communication although not vital could develop in the first phase, and become vital in the subsequent phases where random decisions would prove lethal to the cells.

6.4 Conclusion on the Emergence of Communication in CoCell3

In this chapter, the ambition of this work is partly becoming a reality. We setup a system using cooperative behaviors at every level of the simulation. The cell decision process is modified to favor the emergence of communication and that is a bias that needs to be removed in further versions of CoCell. Nevertheless, using conditions that are not too demanding for cell survival it was possible to observe the emergence of communication protocols between the cells. The process takes time since in the given conditions it is not required for cell survival but nonetheless, some resources emerge as consensus signals to request other resources. Instead of individual signals to request for individual resources, what appears in the system is a universal signal emitted by cells to request help from their neighbors.

This is already an interesting information. If the organized communication hypothesis has any relevance for this system, these results would suggest that communication starts as a very broad purpose one for all signals. During evolution, more signals are probably added to refine the meaning/function of the first one and new combinations appear between these signals for even more functions. Instead of going from a system with one signal per function and then pruning the signals by using combinations of fewer signals to save resources, this system seems to go from a single signal towards a more diverse set. This does actually have some resemblance with real cellular systems like bacteria. Bacteria, as mentioned in chapter 2, have a basic communication system called quorum sensing that allows various colony wide decisions to be taken. Usually these colonies use very few if not a single signal to coordinate their efforts which is similar to what we observe with our system. But many more simulation experiments are necessary to strengthen these arguments.

The very first step to further explore the emergent communication would be to find conditions that allow the appearance of signals but not the survival of non-communicating cells. This would potentially accelerate the selection of signals and speed up the simulation.

The finding that communication in our system only appeared when using very lax conditions raises a question: Were we unable to find the right conditions for the emergence of an essential communication system or are flexible conditions a prerequisite for the appearance of communication? A communication system requires the development of several key innovations, like specific receptors or resources correlations. This requires keeping seemingly useless subsystems in the genetic patrimony until they coalesce into a useful feature. This is probably difficult to do under stressful conditions where resources cannot be diverted from vital functions. So it might be possible that nice conditions are actually required for the development of complex functionalities and that their potential only reveals itself in adversity. This somehow goes against what we know about natural selection and the fact that what is not useful to survive is eventually discarded. This aspect of the emergence of communication would require more investigation.

Then the influence of several structural choices in the simulated system need to be assessed against the apparition of communication, like the diffusion algorithm or the production action mechanism.

To remove the bias towards intercellular communication from the cellular decision process, several paths could be explored. One of them consists in using processes similar to the ones observed in real cells, that is, controlled catalysis and feedback loops. The possibility for one resource to block or enhance the activity of a cellular action without altering its concentration allows basic yet powerful information processing at the cellular level. Up until now we considered that this would interfere with the cooperative decision process. But it could actually be an efficient way to incorporate information about resource correlations to modulate the actions the cell can perform without hardcoding a signal/response mechanism. This action modulation could be decided during the mutation process, and the modifications would be transmitted from generation to generation when they provide a survival advantage.

Modification of CoCell3 into an unbiased communicating system CoCell4 requires more work and testing. But one can only follow a given number of paths before the time is up. And this number of paths is even further reduced when dealing with a system where cells stop at nothing to survive and circumvent the efforts of their observer (and creator). This is why the story of CoCell will pause here for a short while. The development of CoCell4 could be a perfect opportunity for another PhD thesis or a postdoctoral fellowship, who knows.

CHAPTER 7. CONCLUSION AND PERSPECTIVES

The english version starts page 141.

Dans ce chapitre, nous discutons des résultats des expériences CoCell. Pour chaque expérience, nous présentons les motivations, la portée et nous discutons des résultats obtenus. Puis nous discutons des implications de ces résultats sur le problème des communications cellulaires et enfin nous proposons quelques perspectives sur ce travail.

Conclusion sur les expériences CoCell

Cette thèse est basée sur plusieurs observations du domaine de la recherche en biologie. Premièrement, la biologie est aujourd'hui dominée par le paradigme du tout ADN. C'est-à-dire que les maladies et leurs remèdes sont censés être respectivement expliqués et localisés dans le code. Depuis l'achèvement du projet « Human Genome », les attentes ont été élevées pour obtenir des réponses à de nombreux problèmes biologiques. Les remèdes à la plupart des maladies sont contenus dans l'hélice d'ADN puisque toutes les cibles possibles du médicament sont stockées dans cette mémoire héréditaire. Le problème consiste alors seulement à localiser la bonne protéine dans les quelque 4 milliards de bases ou 20 000 gènes, puis à produire une molécule pour moduler sa fonction. Malheureusement, ce n'est pas aussi simple que cela. Peut-être le serait-il si nous n'étions que des machines compliquées comme une voiture ou un ordinateur. Mais nous sommes des machines complexes formées de multiples couches de systèmes complexes interdépendants, le niveau cellulaire étant l'une de ces couches. Essayer de modifier un seul composant en espérant que cela a un seul impact parfaitement défini sur le fonctionnement de l'ensemble de la machine est au mieux un rêve réductionniste. Cela reste cependant la pratique dominante parce que la pensée holistique en biologie est un défi écrasant que nous ne serons pas en mesure de relever dans un avenir proche, faute de données, de méthodologie et de théorie. Le principal problème de ce monde de l'ADN est qu'il a tendance à confiner la biologie dans un monde pseudo-réductionniste où chaque gène a une fonction précise et étouffe les approches alternatives.

Deuxièmement, bien que l'ADN joue un rôle majeur dans la cellule, ce n'est pas le centre de décision central qui contrôle toutes les fonctions d'une cellule, ni ne contrôle les changements qui se produisent ou quand ils se produisent. L'ADN fait partie d'un tout qui est la cellule et les décisions dans ce type de système sont une propriété émergente. Dans ce contexte, quels sont les leviers disponibles pour modifier le comportement cellulaire ? Si nous mettons de côté l'ADN qui est déjà le principal centre d'intérêt de toute l'industrie pharmaceutique, que reste-t-il qui pourrait jouer un rôle important dans la fonction cellulaire ? L'une des pierres angulaires des organismes multicellulaires est la capacité de coopérer au niveau cellulaire. Cette coopération et toutes ses implications sur le fonctionnement des cellules est l'avantage adaptatif qui rend ces types d'organismes possibles. Et comment la coopération est-elle possible pour ces entités chimiques ? Par la division du travail et l'échange d'informations chimiques. Cette information aide à coordonner les actions des groupes de cellules et leur permet de s'entraider au besoin. Ainsi, ce système de communication est très important pour le bon fonctionnement d'une cellule et son altération peut avoir des conséquences désastreuses pour l'ensemble de l'organisme. Ce système de communication est-il basique ou compliqué ? La signification d'un signal est-elle modulée par d'autres signaux reçus par la cellule et par son état interne ? En d'autres termes, s'agit-il de signaux chimiques qui ne

prennent tout leur sens qu'en phrases complètes ? Ce sont des questions auxquelles on ne répond pas entièrement, mais qui pourraient avoir des applications intéressantes pour le développement de médicaments.

Troisièmement, les expériences en laboratoire sur la communication cellule-cellule sont très difficiles et deviennent rapidement impossibles si l'on essaie de traiter des combinaisons de signaux. Pour ces raisons, il n'est pas surprenant que la littérature sur la communication multi-signal soit rare et que les articles sur la communication se concentrent généralement sur l'effet d'un signal unique dans une configuration particulière (généralement pathologique). Une alternative aux expériences du monde réel est la simulation. Ce n'est pas toujours possible, mais lorsque des données brutes sur le système à l'étude, les méthodologies et la puissance de calcul sont disponibles, cela devient un outil attrayant pour étudier un sujet.

L'ensemble de ces observations a conduit au développement de CoCell, un système multi-agent adaptatif (AMAS) pour étudier l'émergence de la communication et sa structure, dans un système multicellulaire simulé simplifié.

Bien qu'il existe de nombreuses méthodes adaptées à la simulation de systèmes biologiques, la plupart d'entre elles nécessitent une connaissance approfondie du fonctionnement interne des cellules et des données expérimentales pour optimiser les différents paramètres. De plus, dans beaucoup de ces méthodes, la convergence est liée au calcul d'une fonction de coût pour l'ensemble du système. Cette fonction de coût peut être très difficile à formaliser et à évaluer pour un système complexe. De plus, cette fonction introduit généralement un biais vers l'état du système correspondant au minimum ou au maximum de la fonction. Habituellement, ce biais est souhaitable puisque dans de nombreuses simulations biologiques, il est souhaité que le comportement du système converge vers les données biologiques réelles. Dans notre cas cependant, un biais d'observation de l'émergence de la communication masquerait les conditions nécessaires à ce phénomène. En d'autres termes, l'utilisation d'une fonction d'évaluation ne peut être utilisée pour essayer d'observer un comportement émergent du système, sauf s'il peut être prouvé que cette fonction n'a pas d'impact sur cette émergence, ce qui peut être extrêmement difficile.

Le choix de la technologie AMAS a été motivé par plusieurs raisons. L'approche AMAS est basée sur la coopération locale entre agents et ne dépend donc pas d'une fonction d'évaluation globale pour converger. Si la coopération est atteinte à tous les niveaux du système, elle devrait produire le comportement global attendu. La coopération est également une caractéristique attendue dans les cellules vivantes, puisque l'interdépendance est essentielle dans les organismes multicellulaires. Le concept de coopération dans les cellules et dans les AMAS n'est peut-être pas exactement de la même nature, mais il rend néanmoins cette approche très bien adaptée à leur étude. Si elle est faite correctement, une mise en œuvre appliquant ce principe peut éviter tout biais en faveur d'un état particulier du système. En fixant les règles de coopération, les raisons de l'émergence de la communication peuvent donc être étudiées au niveau cellulaire local. Enfin, la vitesse de l'exploration spatiale des paramètres peut être considérablement augmentée grâce à la coopération des agents, ce qui est indispensable lorsque le temps et la puissance de calcul sont limités.

Un système AMAS n'est pas incompatible avec d'autres modèles utilisés dans les simulations biologiques comme l'ODE ou les simulations précises de réactions chimiques. Mais nous choisissons d'utiliser des algorithmes simples pour gagner du temps de calcul et pour ajouter à la généricité du système simulé. Comme la communication n'est pas un phénomène observé uniquement pour les cellules, cela signifie qu'elle doit apparaître si les bonnes conditions sont réunies, quel que soit le niveau de détail de la simulation.

Nous avons présenté trois étapes du développement de CoCell. Dans CoCell1, nous avons étudié les règles de coopération de base nécessaires pour s'assurer que chaque cellule sélectionne la bonne action à effectuer au cours de chaque cycle de simulation. De plus, nous avons évalué plusieurs algorithmes pour remplacer efficacement les cellules mortes par des voisins plus efficaces : la pépinière. Seule la pépinière coopérative est capable de préserver un équilibre dynamique dans un système composé de multiples types de cellules interdépendantes et d'assurer sa survie. Diverses propriétés de la pépinière ont été étudiées comme sa robustesse ou son évolutivité. Le système s'est avéré suffisamment robuste pour rester stable même lorsque des choix aléatoires remplacent la prise de décision coopérative. De plus, le processus coopératif était capable de réguler de grands systèmes avec plus de 30 types de cellules interdépendantes dont la taille ne semblait limitée que par la puissance de calcul disponible.

Si l'on oublie un instant que le système est composé de pseudo-cellules et qu'on le considère comme un ensemble de petites usines qui produisent des marchandises à partir de ressources brutes ou des marchandises provenant d'autres usines, les résultats obtenus avec CoCell1 sont intéressants pour l'optimisation des processus et la coordination délocalisée d'entités interdépendantes.

Dans CoCell2, des mutations sont introduites qui permettent à un système de générer la nouveauté nécessaire pour s'adapter à un environnement changeant ou pour devenir plus efficace. Ces mutations sont basées sur des corrélations de ressources afin de former des réseaux d'actions interdépendantes entre les cellules. Ces corrélations correspondent à la base choisie pour l'apparition de la communication dans notre système. Dans une situation où la production d'une ressource est fortement corrélée à l'apparition d'une ressource vitale dans l'environnement, les mutations corrélées s'avèrent efficaces pour maintenir les bonnes actions de production dans la lignée cellulaire et favoriser leur propagation. Nous avons également démontré avec un modèle de l'évolution du système que la propagation des bons « gènes » est liée à la qualité des corrélations suivant une courbe quasi exponentielle. Cela signifie que seules des corrélations presque parfaites sont capables de propager à l'ensemble du système les actions de production essentielles. Comme il est presque impossible d'identifier toutes les corrélations pertinentes dans l'environnement et de rejeter tous les faux positifs possibles, ce modèle a montré qu'il serait illusoire d'espérer un système parfaitement stable avec une population cellulaire homogène. D'autre part, l'hétérogénéité est toujours une bonne source de nouveauté et de comportements intéressants.

Enfin, CoCell3 étudie l'émergence de la communication intercellulaire. Tout d'abord, en utilisant des contraintes sur le comportement cellulaire et un biais vers la communication dans le mécanisme de décision cellulaire, le système est testé pour évaluer sa capacité à évoluer vers l'échange d'informations. Dans une configuration où les conditions rendent la survie des cellules assez facile, une certaine forme de communication est observée entre les cellules du système. Un signal unique semble être utilisé par toutes les cellules pour demander aux cellules voisines ce qui leur manque pour survivre. Aucune combinaison de messages n'est détectée qui pourrait être utilisée pour une demande spécifique, mais cela aurait été une surprise puisque ce premier système n'a pas été conçu pour l'émergence de phrases de signalisation. Bien que la communication soit présente, elle n'est pas essentielle à la survie du système car les conditions sont très indulgentes. Mais chaque fois que les circonstances de départ sont plus exigeantes pour les cellules, le système ne peut pas survivre assez longtemps pour développer des compétences de base en communication. Ces conditions douces pourraient être une condition préalable à l'émergence de caractéristiques compliquées comme la communication dans un système qui lutte pour survivre. Pour en être sûr, davantage d'expériences doivent être réalisées et divers algorithmes coopératifs testés.

Arriver à la fin du temps alloué à la thèse à cette étape de la recherche lorsque des résultats intéressants commencent à émerger est toujours frustrant. Mais cela ne signifie pas que le travail

doive s'arrêter là et un jour la vérité sur le langage secret des cellules sera dévoilée. Et ce jour, je suis sûr que la médecine fera un grand pas en avant.

Pendant deux ans, j'ai réalisé plusieurs centaines de simulations avec CoCell. Nous pouvons en tirer trois grandes leçons :

- *(i) dans cet espace gigantesque de paramètres globaux du système, très peu d'entre eux conduisent à des états dynamiques « solutions » stables et dynamiques ;*
- *(ii) lorsque de telles solutions existent, le processus décisionnel coopératif local semble être en mesure de les trouver dans un délai de calcul raisonnable ;*
- *(iii) parmi ces états de solution, très peu imposent la communication intercellulaire pour durer. Cela ne signifie pas qu'une communication explicite est inutile. Mais les conditions de son émergence nécessitent probablement que l'environnement ne soit pas trop bruité (les ressources fluctuent trop ou sont proches du niveau de saturation) pour que des corrélations constructives puissent être établies.*

Cette analyse peut avoir des conséquences sur les systèmes auto-organisés artificiels tels que ceux basés sur la théorie AMAS. En effet, l'auto-organisation conduit parfois les agents à s'engager dans l'apoptose parce qu'ils se considèrent comme inutiles au sein du collectif. Avant l'apoptose, ils pourraient essayer de faire émerger de nouveaux « protocoles » de communication dans le réseau d'agents qui donneraient des capacités de survie supplémentaires au système et élargiraient ainsi les moyens d'adaptation collective.

Dans CoCell, par conception, les agents cellulaires travaillent avec le moins d'informations possible. Non seulement un agent n'a qu'une vision très partielle de son environnement, mais il ignore aussi que d'autres agents sont situés dans son voisinage (il suppose simplement qu'ils doivent exister parce que les ressources fluctuent, et qu'ils ont le même comportement coopératif parce que c'est la base de la théorie AMAS). Ce travail est un test pour les AMAS parce que de telles situations n'ont pas souvent été traitées. Même s'ils ne peuvent pas communiquer directement, les robots de transport (Picard 2004) sont capables de distinguer les robots et les boîtes à transporter ou de savoir où se dirige un autre robot. Même si elles communiquent indirectement, les fourmis (Topin et al. 1999) déposent de la phéromone parce qu'elles savent que cette phéromone signifiera quelque chose de spécial pour les autres fourmis, ce qui n'est pas le cas ici lorsque les cellules libèrent des ressources dans l'environnement.

Habituellement, l'attitude coopérative des agents peuplant un AMAS est basée sur un comportement dit nominal, où un agent, s'il est pleinement coopératif ne devrait jamais être confronté à des « Situations Non Coopératives » (SNC), et un comportement dit « coopératif » qui tente d'éviter ou de réparer d'éventuels SNC. Dans CoCell, l'attitude coopérative des agents est construite différemment, en la considérant comme une combinaison de comportements égoïstes et altruistes, respectant le fait qu'un agent coopératif n'est pas totalement altruiste mais est bienveillant parce qu'il agit de manière à ne pas se sacrifier pour les autres en évitant de devenir le plus critique. Il s'agit d'une approche différente de la conception du comportement local qui pourrait être utile dans d'autres systèmes basés sur les AMAS.

Perspectives

CoCell n'est qu'un point de départ pour explorer plus avant l'évolution de la communication. De nombreuses hypothèses doivent encore être mises en œuvre et testées. De nombreux scénarios et configurations de systèmes pourraient également être explorés.

La première étape à suivre dans le travail de CoCell est d'éliminer le biais de communication introduit dans CoCell3 et de répéter les expériences pour lesquelles la communication a été

observée. Différentes méthodes sont possibles pour éliminer le biais dans le processus de prise de décision des cellules. L'une d'entre elles consiste à introduire des boucles de rétroaction : les actions cellulaires ne pourraient être utilisées que lorsque la concentration d'une ressource donnée est supérieure ou inférieure à un seuil. Un autre, plus difficile à mettre en œuvre, serait de permettre l'évolution du processus décisionnel lui-même par le biais de mutations. Si dans le système non biaisé, la communication émerge toujours, alors d'autres mesures peuvent être prises afin de prouver ou de réfuter l'hypothèse de la communication par mot/phrase. Une étape nécessaire serait de rendre les fonctions cellulaires plus élaborées afin que les cellules dépendent de plus de types de ressources que la seule énergie. En effet, il serait difficile d'observer divers signaux liés à différents besoins cellulaires essentiels s'il n'y a qu'une ou deux ressources nécessaires à la survie cellulaire. De plus, ces besoins cellulaires devraient avoir différents degrés de criticité.

La notion de parenté dans le développement de la multicellularité, donc de la communication, doit être étudiée plus avant. D'après la littérature, il semble que la similarité du profil cellulaire est la clé pour construire un organisme multicellulaire fort. Dans nos expériences, nous avons utilisé la parenté pour concevoir certains aspects des comportements coopératifs. Il serait intéressant d'y donner suite en simulant un système composé de plusieurs colonies de cellules non apparentées se disputant le même espace. Certaines colonies pourraient être composées de cellules fortement similaires tandis que d'autres pourraient avoir des liens de parenté faibles. Dans ces conditions, l'influence de la parenté sur l'évolution multicellulaire et l'émergence de la communication pourrait être évaluée.

En concevant CoCell, nous avons décidé très tôt d'avoir des agents cellulaires très génériques et extrêmement simplifiés. Nous avons fait ce choix principalement pour deux raisons :

- *Premièrement, notre compréhension actuelle des mécanismes présents dans une cellule vivante est encore limitée et il est difficile d'essayer de les reproduire avec précision.*
- *Deuxièmement, le concept de communication ne semble pas être spécifique aux cellules vivantes. Elle est apparue dans d'autres systèmes comme le cerveau humain.*

Comprendre ou reproduire les mécanismes internes d'une cellule vivante sont au centre des préoccupations de nombreuses équipes de recherche dans le monde entier. De plus, même avec une bonne représentation du fonctionnement interne de la cellule, la définition d'un point de départ pour la simulation reste très difficile. Nous travaillons sur des systèmes intrinsèquement chaotiques et les conditions initiales sont très importantes pour leur évolution et leur stabilité. Il est déjà si difficile de trouver des conditions de démarrage pertinentes dans un système aussi simple que CoCell que ce serait probablement un cauchemar pour une simulation cellulaire même légèrement plus précise. Évidemment, des données réelles sur l'état d'une cellule résoudraient ce problème, mais pour l'instant, ces expériences sont encore hors de portée.

L'émergence de la communication n'est pas spécifique aux cellules vivantes et cela signifie peut-être que ce phénomène n'est pas spécifique à un ensemble de conditions et de mécanismes mais plutôt à un ensemble de fonctionnalités et d'organisations. C'est-à-dire que, même si les lois de la chimie et de la physique étaient différentes, si une collection d'entités peut effectuer certaines transformations sur leur environnement et partager certaines informations, la communication émergerait quand même. Dans cette optique, nous ne voulions pas seulement étudier l'avènement de la communication dans les communautés cellulaires, mais aussi déterminer les caractéristiques nécessaires et suffisantes pour que ce phénomène apparaisse. Qui sait, puisque les étoiles sont des entités qui transforment leur environnement et partagent des ressources, elles communiquent peut-être sur une échelle de temps de plusieurs millions d'années. Ou peut-être pas.... Parce que d'autres caractéristiques sont probablement nécessaires comme l'interdépendance ou d'autres

caractéristiques non encore identifiées. Cependant, si cette hypothèse est vraie, la pertinence physique et chimique et la précision de la simulation ne devraient pas avoir un impact majeur sur le résultat tant que toutes les caractéristiques cruciales requises pour la communication sont incluses. De plus, les simplifications rendent l'analyse du système plus aisée, facilitent le développement du code et réduisent le temps de calcul.

Comme nous l'avons mentionné dans le chapitre introduisant le contexte de ce travail, le véritable test pour toute théorie ou hypothèse est l'expérience réelle. L'inconvénient de notre approche est que la généricité du système et les simplifications rendent difficile la validation des résultats dans des systèmes réels.

Afin de tester en laboratoire une partie de l'hypothèse générée avec CoCell, il est probable que le système doit d'abord se rapprocher des cellules réelles dans les algorithmes utilisés et dans le niveau de détail des différents processus. Plusieurs simplifications effectuées dans CoCell doivent être supprimées et remplacées par des descriptions plus précises des processus cellulaires. En particulier, il devrait être possible d'intégrer dans la simulation des données issues d'expériences réelles. De cette façon, le résultat de la simulation pourrait être comparé avec le comportement cellulaire observé.

Enfin, plusieurs choix structurels faits dans le développement de CoCell pourraient également être contestés. En particulier, la taille cellulaire constante peut avoir un impact sur le comportement à long terme du système et il serait intéressant d'évaluer son influence dans une version modifiée où les cellules peuvent augmenter ou diminuer leur taille. Évidemment, cela aurait aussi des conséquences sur la taille des nœuds et le voisinage de chaque cellule. Un modèle sans grille pourrait être intéressant à utiliser dans ces conditions.

Toutes ces extensions du modèle CoCell nécessiteront du temps et des tests, et elles devraient être classées par ordre de priorité en fonction de leur impact potentiel sur la démonstration de l'hypothèse de communication organisée.

In this chapter, we discuss the results of the CoCell experiments. For each experiment, we present the motivations, the scope and we discuss the results obtained. Then we discuss the implications of these results on the cellular communication problem and finally we propose some perspectives about this work.

7.1 Conclusion on the CoCell Experiments

This thesis is based on several observations of the biological research field. Firstly, biology is dominated nowadays by the all-DNA paradigm. That is, diseases and their cures are supposedly respectively explained and located in the code. Since the completion of the Human Genome Project, expectations have been high to get answers to many biological problems. Cures to most diseases are contained in the DNA helix since all possible drug targets are stored in this hereditary memory. The problem then only consists in pinpointing the right protein in the nearly 4 billion bases or 20000 genes and then produce a molecule to modulate its function. Unfortunately, it is not as simple as that. Maybe it would be if we only were complicated machines like a car or a computer. But we are complex machines formed of multiple layers of interdependent complex systems, the cellular level being one of these layers. Trying to alter a single component hoping that this has a single perfectly defined impact on the workings of the whole machine is at best a reductionist dream. This is the way it is done because thinking holistically in biology is an overwhelming challenge that we will not realistically be able to address in the near future for lack of data, methodology, and theory. The main problem with this DNA world is that it has a tendency to confine biology in a pseudo-reductionist world where each gene has a precise function.

Secondly, although DNA has a major role in the cell, it is not the central decision center that controls every function that a cell does, nor does it control which changes occur or when. DNA is part of a whole that is the cell and decisions in this type of system are an emergent property. In that context, what are the levers available to modify the cell behavior? If we put aside DNA that is already the main focus of all the pharmaceutical industry, what is left that could play an important role in the cell function? One of multicellular organisms' cornerstones is the ability to cooperate at the cellular level. This cooperation and all its implications on cell workings is the fitness bonus that makes these types of organisms possible. And how is cooperation possible for these chemical entities? Through division of labor and the exchange of chemical information. This information helps coordinate the actions of groups of cells and allows them to help each other when required. Thus, this communication system is very important for the smooth operation of a cell and its alteration can have dire consequences for the whole organism. Is this communication system basic or complicated? Is the meaning of a signal modulated by other signals received by the cell and by its internal status? In other words, are the chemical signals words that only take their full meaning in full sentences? These are questions that are not fully answered but could have interesting applications for drug development.

Thirdly, wet lab experiments on cell-cell communication are very difficult and quickly become intractable if we try to deal with signal combinations. For these reasons it is not surprising that literature on multi-signal communication is rare and that papers on communication are usually focused on the effect of a single signal in a particular setup (usually pathological). An alternative to real world experiments is simulation. This is not always possible, but when raw data about the system under investigation, methodologies and computing power are available, this becomes an attractive tool to investigate a topic.

Together, these observations lead to the development of CoCell, an AMAS system to study the emergence of communication and its structure, in a simple simulated multicellular system.

Although there are numerous methods suited to simulate biological systems, most of them require in depth knowledge of the inner workings of cells and experimental data to tune the various parameters. Furthermore, in many of these methods the convergence is tied to the calculation of a fitness function for the whole system. This evaluation function can be very difficult to formalize and to evaluate for a complex system. Also, more often than not this function introduces a bias towards the state of the system corresponding to the minimum or the maximum of the function. Usually, this bias is desirable since in many biological simulations, we want the system behavior to converge towards the actual biological data. In our case though, a bias to observe the emergence of communication would mask the necessary conditions for this phenomenon. In other words, the use of an evaluation function cannot be used when trying to observe an emerging behavior of the system unless it can be proven that the function does not have any impact on this emergence, and this can be extremely difficult.

The choice of the AMAS approach was motivated by several aspects. The AMAS methodology is based on local cooperation between agents and does not depend on a global evaluation function to converge. If cooperation is reached at every level of the system, it should produce the expected global behavior. Cooperation is also an expected feature in living cells, since interdependency is key in multicellular organisms. The cooperation concept in cells and in AMAS might not be exactly of the same nature, nonetheless it makes this methodology very well suited for their study. If done correctly, an implementation following this principle can avoid any bias towards a special state of the system. Setting the cooperation rules, the reasons for the emergence of communication can therefore be studied at the local cellular level. Finally, the speed of the parameter space exploration can be greatly increased using agent cooperation and this is dearly needed when only limited time and computing power are available.

An AMAS system is not incompatible with other models used in biological simulations like ODE or precise chemical reaction simulations. But we choose to use simple algorithms to save computation time and to add to the genericity of the system being simulated. Since communication is not a phenomenon only observed for cells, it means that it should appear if the right conditions are present, irrespective of the level of details of the simulation.

We presented three stages of the development of CoCell. In CoCell1, we studied the basic cooperation rules needed to ensure that each cell selects the right action to perform during each simulation cycle. Also, we evaluated several algorithms to replace efficiently dead cells by more efficient neighbors: the nursery. Among the tested algorithms, only the cooperative nursery is able to preserve a dynamic equilibrium in a system composed of multiple interdependent cell types and its continuous survival. Various properties of the nursery have been investigated like its robustness or scalability. The system proved to be robust enough to stay stable even when random noise replaced cooperative decision-making most of the time. Also, the cooperative process was capable to regulate large systems with more than 30 interdependent cell types with a size seemingly only limited by the computing power available.

If for an instant we forget that the system is composed of pseudo-cells and consider it as an ensemble of small factories that produce goods from raw resources or goods from other factories, the results found with CoCell1 are of interest for process optimization and delocalized coordination of interdependent entities.

In CoCell2, mutations are introduced that allow a system to generate the novelty necessary to adapt to a changing environment or to become more efficient. These mutations are based on resource correlations in order to form networks of interdependent actions between the cells. These correlations correspond to the chosen basis for the apparition of communication in our system. In a situation where production of a resource is strongly correlated to the apparition of a vital resource

in the environment, correlation-based mutations prove to be efficient in keeping the right production actions in the cell line and promote their propagation. We also demonstrated with a model of the system evolution that the propagation of the right "genes" is tied to the quality of the correlations following a near exponential curve. That means that only near perfect correlations are able to propagate to the whole system the essential production actions. Since it is near impossible to identify all relevant correlations in the environment and discard all possible false positives, this model showed that it would be illusory to hope for a perfectly stable system with a homogenous cellular population. On the other hand, heterogeneity is always a good source of novelty and interesting behaviors.

Finally, in CoCell3 the actual emergence of intercellular communication is investigated. First, using constraints on the cell behavior and a bias towards communication in the cell decision mechanism, the system is tested to evaluate its capacity to evolve information exchange. In a setup where conditions make cell survival quite easy, some form of communication is observed between the cells of the system. A one-size-fits-all signal appears to be used by all cells to request from neighboring cells what they lack to survive. No message combination is detected that could be used for specific request, but that would have been a surprise since this first system was not designed for the emergence of signal sentences. Although, communication is present, it is not essential for the survival of the system because conditions are quite forgiving. But whenever starting circumstances are more demanding on cells, the system cannot survive long enough to develop basic communication skills. These mild conditions might be a prerequisite for the emergence of complicated features like communication in a system struggling to survive. To be sure, more experiments need to be performed and diverse cooperative algorithms tested.

Reaching the end of the allotted time at this stage of the research where interesting results start to emerge is always frustrating. But it does not mean the work must stop there and one day the truth about the secret language of cells will be unveiled. And this day I am sure medicine will take a big step forward.

For two years I performed several hundred simulations with CoCell. We can draw three main lessons from this:

(i) in this gigantic space of global system parameters, very few lead to stable dynamic "solution" states;

(ii) when such solutions exist, the local cooperative decision-making process seems to be able to find them in a reasonable computational time;

(iii) among these solution states, very few impose intercellular communication to last. This does not mean at all that the need for explicit communication is practically useless. But the conditions for its emergence probably require that the environment is not too noisy (resources fluctuate too much or are near saturation level) so that constructive correlations can be established.

This analysis can have consequences on artificial self-organizing systems such as those based on the AMAS theory. Indeed, self-organization sometimes leads agents to engage in apoptosis because they consider themselves useless within the collective. Before apoptosis, they could try to make new "protocols" of communication emerge in the network of agents which would give additional capacities of survival to the system and thus widen the means of collective adaptation.

In CoCell, by design the cell agents work with as little information as possible. Not only does an agent has only a very partial view of its environment but it also ignores that other agents are located in its neighborhood (it just assumes that it must exist because resources fluctuate, it assumes that others have the same cooperative behavior because that is the basis of the AMAS theory). This

work is a test for AMAS because such situations have not often been dealt with. Even if they cannot communicate directly, transport robots in (Picard 2004) are able to distinguish robots and boxes to be transported or to know where another robot is headed. Even if they communicate indirectly, ants in (Topin et al. 1999) deposit pheromone because they know that this pheromone will mean something special to other ants, which is not the case here when cells release resources into the environment.

Usually, the cooperative attitude of agents populating an AMAS is based on a so-called nominal behavior, where an agent if fully cooperative and should never be confronted with "Non Cooperative Situations" (NCS), and a so-called "cooperative" behavior which tries to avoid or repair possible NCS. In CoCell, the cooperative attitude of agents is constructed differently, seeing it as a combination of selfish behaviors and altruistic behaviors, respecting the fact that a cooperative agent is not totally altruistic but is benevolent because it acts in such a way as not to sacrifice itself for others by avoiding becoming the most critical. This is a different approach to the design of local behavior that could be useful in other systems based on AMAS.

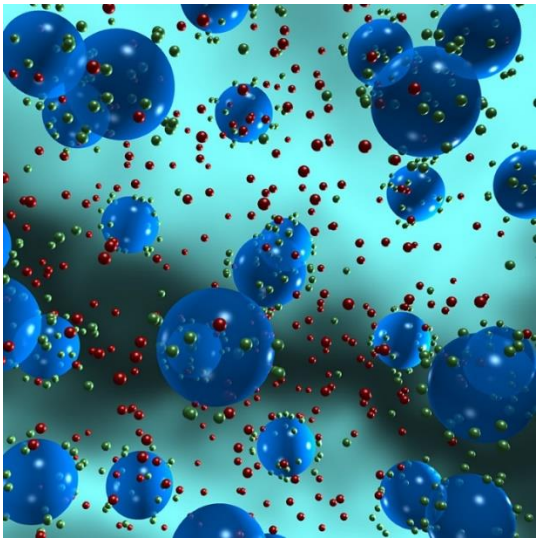


Figure 7-1 Artist view of cells that communicate by exchanging molecules. Purdue University. Credit: Ken Ritchie.

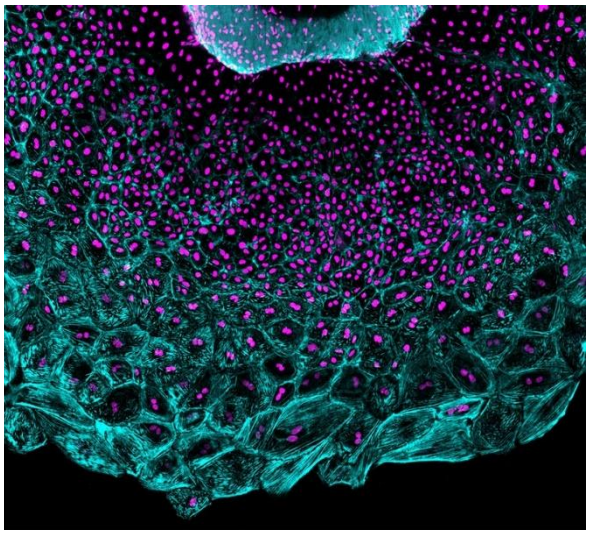


Figure 7-2 Regrowing cells in a zebrafish heart coordinated by communication. Credit: Jingli Cao, Duke University

CoCell is but a starting point to further explore the evolution of communication. Many hypotheses still need to be implemented and tested. And many scenarios and system configurations could be explored as well.

The first step to follow on the CoCell work is to remove the bias towards communication introduced in CoCell3 and repeat the experiments where communication was observed. Different ways are possible to remove the bias in the cell decision-making process. One of them is to introduce feedback loops: cell actions can only be used when the concentration of a given resource is above/below a threshold. Another one, more difficult to implement, would be to enable the evolution of the decision process itself through mutations. If in the unbiased system, communication still emerges then further steps can be taken in order to prove or disprove the word/sentence communication hypothesis. One necessary step would be to render the cell functions more elaborated so that cells are dependent on more things than just energy. Indeed, it would be difficult to observe various signals linked to different essential cell requirements if there are only one or two required resources for cell survival. Also, these cellular needs should have different degrees of criticality.

The notion of kinship in the development of multicellularity, hence communication, needs to be investigated further. From the literature it appears that cell profile similarity is key to build a strong multicellular organism. In our experiments, we used kinship to design some aspects of the cooperative behaviors. It would be interesting to follow up on this by simulating a system composed of several unrelated cell colonies competing for the same space. Some colonies could be composed of strongly similar cells while others could have loose relatedness. In these conditions, kinship influence on multicellular evolution and communication emergence could be assessed.

Designing CoCell, we decided very early on to have cell agents that are very generic and extremely simplified. This was our choice mainly for two reasons:

- Firstly, our current understanding of the mechanisms present in a living cell are still limited and trying to replicate them accurately is difficult.
- Secondly, the concept of communication does not appear to be specific to living cells. It emerged in other systems like human brains.

Understanding or replicating the inner mechanisms of a living cell are the focus of many research teams worldwide. Furthermore, even with a good representation of the inner workings of the cell, the definition of a starting point for the simulation is still very challenging. We are working on intrinsically chaotic systems and initial conditions are very important for their evolution and stability. It is already so difficult to find relevant starting conditions in a system as simple as CoCell that it would probably be a nightmare for an even slightly accurate cellular simulation. Obviously, actual data on the state of a cell would solve this issue but for now these experiments are still beyond reach.

Communication emergence is not specific to living cells and this possibly means that this phenomenon is not specific to a set of conditions and mechanisms but more on a set of functionalities and organizations. That is, even if the laws of chemistry and physics were different, if a collection of entities can perform some transformations on their environment and share some kind of information, communication would still emerge. With this in mind we not only wanted to study the advent of communication in cell communities but also determine the necessary and sufficient features required for this phenomenon to appear. Who knows, since stars are entities that transform their environment and share resources, maybe they are communicating on a time scale of millions of years. But maybe not… Because some other features are probably required like interdependency or other still unidentified characteristic. Still, if this hypothesis is true, the physical and chemical relevance and accuracy of the simulation should not have a major impact on the outcome as long as all the crucial features required for communication are included. Moreover, simplifications render the analysis of the system more tractable, the development of the code easier, and reduces the computation time.

As mentioned in the context chapter, the true test for any theory or hypothesis is actual experiments. The drawback of our approach is that the genericity of the system and the simplifications make it difficult to readily validate the results in real-life systems.

In order to test in the wet lab some of the hypothesis generated with CoCell, it is probable that the system first needs to get closer to real cells in the algorithms used and in the level of details for the various processes. Several simplifications that are made in CoCell need to be removed and replaced by more precise descriptions of the cellular processes. In particular, it should be possible to integrate data from real experiments in the simulation. In this way the outcome of the simulation could be compared with observed cellular behavior.

Finally, several structural choices made in the development of CoCell could also be challenged. In particular, the constant cellular size may have an impact on long term system behavior and it

would be interesting to assert its influence in a modified version where cells can grow or shrink their size. Obviously this would also have consequences on the node size and neighborhood of each cell. A lattice-free model might be interesting to use in these conditions.

All these extensions of the CoCell model will require time and testing, and they should be prioritized by their potential impact on the demonstration of the organized communication hypothesis.

ANNEX – COCELL WITH BUILT-IN COMMUNICATION

The first six months of the thesis were devoted to the design of a system where communication between cells already exists. The goal of this system was to show that it is possible to differentiate between two types of communication – one signal for one response and word/sentence – by their impact on the survival capacity of the system as a whole and on its energy balance.

Definitions

The system is designed as a MAS and not an AMAS: the (scripted) behavior of a cell agent is not necessarily cooperative, and there is no adaptation. Several agent types can be present in the simulation to describe various cell types, each one with its own set of scripts. Mutations were not yet included when the switch to the system finally adopted for CoCell occurred.

Most characteristics defining the cellular agents and the environment are similar to CoCell:

- A 2D or 3D toroid space.
- Each node of this space contains a cell agent.
- Nodes and cells have an arbitrary set of neighbors (usually 8 in 2D and 26 in 3D).
- A set of resources is present on each node and in each cell.
- Resources present on nodes circulate in the system following a passive diffusion mechanism.

Nevertheless, there are several important differences, notably for the cell properties and its behavior:

- There are signals which are distinct from resources.
- As resources, signals circulate following a passive diffusion mechanism. Furthermore, they degrade with time when released in the environment thus limiting the communication range.
- A cell is defined by:
    - A type and a position;
    - Its total Energy: E_{tot};
    - The Resources in stock: $\{R_0, \ldots, R_{nm}\}$;
    - The Signals in stock: $\{s_0, \ldots, s_{ns}\}$;
    - Expressed receptors: $\{r_0, \ldots, r_{nr}\}$. Receptors present at the cell surface represent its capacity to register the presence of specific signals. Receptor r_i detects signal s_i;
    - A set of actions.
- The common actions shared by all cells are the following:
    - Maintain E_{tot} between E_{Min} and E_{Max}:
        - By gathering resources: E- (consumes energy)
        - By transforming resources into energy: E+ (produces energy)
    - Maintain stock of resources to a nominal value (between R_i^{Min} and R_i^{Max}):

- By gathering from environment: E-
- By requesting from neighbor cells: sending messages E-

- Maintain receptor levels (between r_i^{Min} and r_i^{Max}):
  - By producing new receptors according to cell type specifications: E-
  - The energy cost per receptor is proportional to the number of types of receptors to maintain
    - This accounts for the maintenance of a larger internal "machinery"
- Maintain stock of signals (between s_i^{Min} and s_i^{Max}):
  - By producing new signals according to cell type specifications: E-
  - The energy cost per signal is proportional to the number of types of signals to maintain
    - This accounts for the maintenance of a larger internal "machinery"
- Receive signals via receptors:
  - When the product between the concentration of a signal in the environment and its specific receptor is above a given threshold, the cell registers the signal:

    $[s_i] \times [r_i] > Sensitivity_i$ then s_i is detected *i.e.* Si = true
  - Stock of specific receptor decreases because of endocytosis (internalization and degradation): E-
  - Sensitivity to this signal decreases until receptor concentration is restored
  - Reception of a signal can generate a cellular response like a simple production action, cell division or apoptosis (depending on the cell script)
- Emit signals when required by the cell script behavior:
  - Send signals in the environment: E-
  - Stock of signals decreases
  - A signal can be detected by the cell that emitted it if it has the right receptor

The cell scripts are very simple. They consist of formulas and recipes.

- A formula describes a set of signals detection necessary to activate a given cell action:

  $Formula$: $Given\ a\ set\ ID\ \{id_1, \dots, id_n\}, if\ \sum_{i\ \in\ ID} S_{id_i} = n\ then\ execute\ action_{formula}$
- A recipe describes a production action:

  Recipe: Given the following sets

  $ID^{reac}\ \{reac_1, \dots, reac_n\}, ID^{cond}\ \{cond_1, \dots, cond_m\}, ID^{prod}\ \{prod_1, \dots, prod_p\}$

  the production action is

  $$\sum_{i\ \in\ ID^{reac}} a_i * R_i \xrightarrow{\sum_{k\ \in\ ID^{cond}} Test_k = m} \sum_{j\ \in\ ID^{prod}} b_j * R_j$$

  Where $\{a_i\}$ and $\{b_i\}$ are stoichiometric factors with $\sum_i a_i = 1\ and \sum_i b_i = 1$

  $Test_i$ is a condition that is equal to 1 if either $\left([R_i] > threshold_i^{min}\right), ([R_i] < threshold_i^{max})$ or $(threshold_i^{min} < [R_i] < threshold_i^{max})$ depending on the recipe.

The case where a cell dies and needs to be replaced was not treated in this system. All test scenarios made sure each cell received enough energy related resources to sustain a continuous operation.

From this definition and rules, the system was tested in several scenarios to evaluate the choice of parameters and the stability of the solutions found.

Scenario Examples

Some examples of working scenarios are described thereafter:

i. Active Diffusion of a Resource Following Signal Requests in a 3D Space

The setup for this scenario is as follows:

- Energy:
  - Cells will transform material into energy according to their energy gain and available quantities
  - When a material is depleted, the cell will try to gather more from the environment
- Messages:
  - Resource R_0 is only present in the center or on the side. It cannot be used for energy
  - Resource R_1 is present in vast quantities everywhere and is used for energy
  - All cells lacking R_0 executes the corresponding message formula to request it

As shown on Figure 3 and Figure 4, cells receiving the right signals release some of the R_0 they have in stock into the environment. R_0 tends to become uniformly distributed after a number of cycles and all cells end up having R_0^{mean} in stock. At this moment, the number of requests drops sharply.

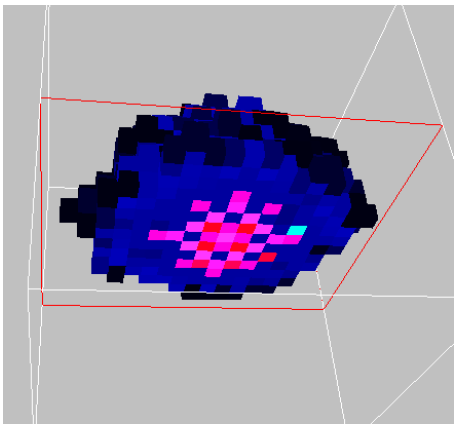


Figure 3 Propagation of R_0 from the center using active diffusion.

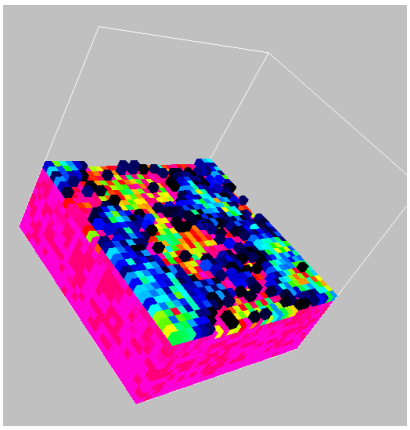


Figure 4 Propagation of R_0 from the bottom of the system. Following requests from cells, R_0 diffuses to the rest of the box.

ii. Chain Production Scenario in 2D

In this scenario, there are three different cell types and one "cell type" that represents empty channels where resources can diffuse passively. The goal is to production R_0 from two other resources R_3 and R_4 synthesized by two cell types (Figure 5).

- Cell type 0 produces R_0 from R_4 and R_3 and can request R_3 and R_4 using signals s_0 and s_1

- Cell type 1 produces R_4 from R_1 and R_2 (both available in large quantities)
- Cell type 2 produces R_3 from R_1 and R_2
- Cell type 3 represent channels that transport R_0 out of the system

Cells are distributed in the system space following a procedural simplex noise in a way that embeds cell types 0 inside a layer of cell types 1 and itself in a layer of cell types 2. In this way the messaging mechanism is necessary to provide cell type 0 with nutrients and to release R_4 in the channels. Cell type 0 needs to request R_3 and R_4 from cell types 1 and 2, and releases R_0 which should be actively transported (using signal/response mechanism) through cell types 1 and 2 until it reaches the channels where it diffuses passively.

Figure 5 on the right shows the concentration of R_4 after the system started producing R_4 and before a dynamic equilibrium is reached where concentration of R_4 becomes homogeneous in the channels. We can see that most of the channels are starting to fill with R_4. Type 0 cells have various levels of R_4 depending on their current production. Levels in type 1 and 2 are not represented for clarity.

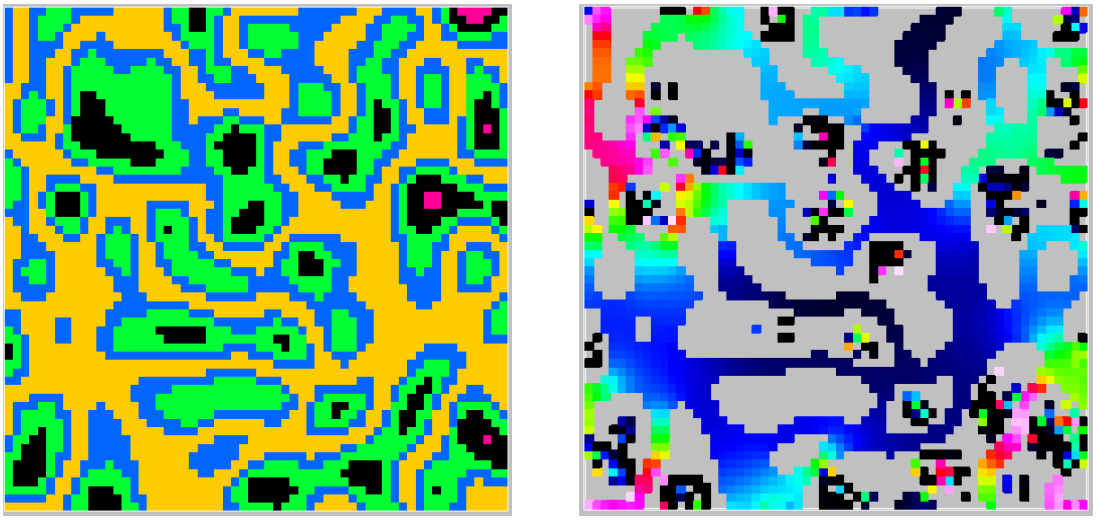


Figure 5 Left: Cell types: black cells produce R_0, green cells produce R_3, blue is for R_4 and orange is for channels. Right: concentration of R_4 diffusing in the channels (from blue to red).

iii. Altruist Cells Scenario

This setup is very similar to (i) where R_0 is initially located only in the center of the system space. The difference is that cells behave altruistically and do not hesitate to yield all their personal stock of R_0 to help neighboring cells if they receive the appropriate request (Signal 0). What happens in this case is that during the phase leading to a dynamic equilibrium, cells alternate between requesting R_0 and releasing it to help others. This creates a wave/pulsating pattern that is difficult to reproduce on paper but is visually very "organic" in shape and evolution (Figure 6). In the end all cells reach their target stock of R_0 and stop sending requests.

The same experiment performed in a 3D space displays the same kind of behavior but at some point all cells become synchronized and alternatively gather and release R_0 together in the environment, creating a pulsating time pattern as illustrated in Figure 7 bottom row.

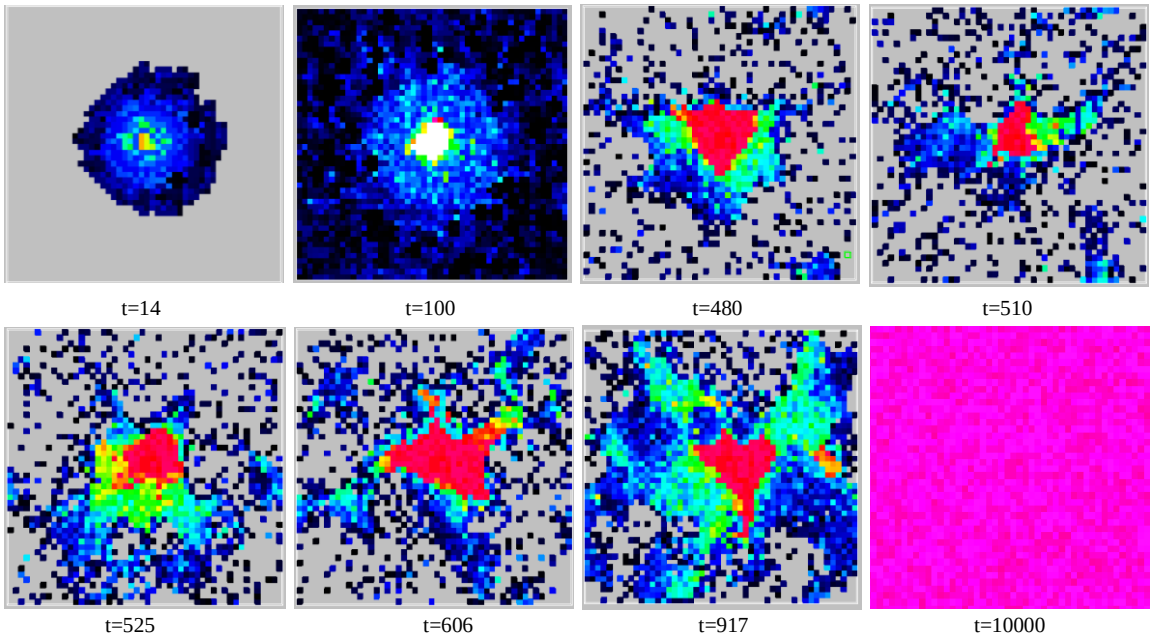


Figure 6 Propagation of R_0 from the center using active diffusion in altruist cells.

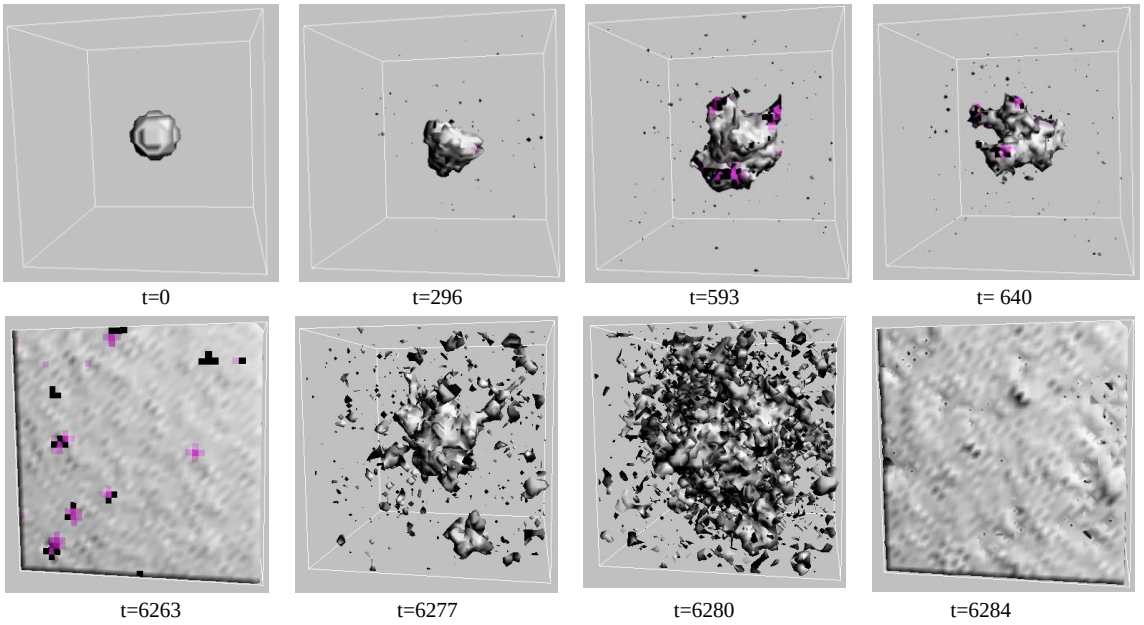


Figure 7 Propagation of R_0 from the center using active diffusion in altruist cells in 3D.

Conclusion

Although some interesting results can be obtained using this system, its potential for development seemed limited. The main problem with the adopted approach is that messages and responses to messages are "hard coded" in the script followed by the cells. In other words, we establish what we want to see in the behavior of the cells in their script. Sometimes, unexpected results are observed like the pulsating patterns, but there is no adaptation of the cells to their environment. Mutations could have improved that situation but it was difficult to implement in this scripting framework. Another problem was the tuning of the various parameters of the system and in the scripts that could induce very strong variations in the results.

BIBLIOGRAPHY

Adams, Douglas. 1980. “The Hitchhiker’s Guide to the Galaxy - UW-Madison Libraries.” *New York :Harmony Books, 1980. Print.*

Adcock, Stewart A, and J Andrew McCammon. 2006. “Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins.” *Chemical reviews* 106(5): 1589–1615.

Aktipis, C Athena et al. 2015. “Cancer across the Tree of Life: Cooperation and Cheating in Multicellularity.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1673).

Alves, Rui, Fernando Antunes, and Armindo Salvador. 2006. “Tools for Kinetic Modeling of Biochemical Networks.” *Nature Biotechnology* 24(6): 667–72.

An, Gary, Qi Mi, Joyeeta Dutta-Moscato, and Yoram Vodovotz. 2009. “Agent-Based Models in Translational Systems Biology.” *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1(2): 159–71.

Andrews, Steven S, Nathan J Addy, Roger Brent, and Adam P Arkin. 2010. “Detailed Simulations of Cell Biology with Smoldyn 2.1.” *PLoS computational biology* 6(3): e1000705.

Aruoba, S Boragan, Jesus Fernandez-Villaverde, and Juan F Rubio-Ramirez. 2006. “Comparing Solution Methods for Dynamic Equilibrium Economies.” *Journal of Economic dynamics and Control* 30(12): 2477–2508.

Baars, Bernard J., and David B. Edelman. 2012. “Consciousness, Biology and Quantum Hypotheses.” *Physics of Life Reviews* 9(3): 285–94.

Ballet, Pascal et al. 2017. “Modelling and Simulating Complex Systems in Biology.” In IGI Global, 128–58.

Barbuti, Roberto et al. 2011. “Spatial P Systems.” *Natural Computing* 10(1): 3–16.

Becker, Scott A et al. 2007. “Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox.” *Nature Protocols* 2(3): 727–38.

Ben-Shlomo, Izhar et al. 2003. “Signaling Receptome: A Genomic and Evolutionary Perspective of Plasma Membrane Receptors Involved in Signal Transduction.” *Science’s STKE : signal transduction knowledge environment* 2003(187): RE9.

Bernado, P, and M Blackledge. 2010. “Proteins in Dynamic Equilibrium.” *Nature* 468(7327): 1046–47.

Bernon, Carole, Marie-Pierre Gleizes, Sylvain Peyruqueou, and Gauthier Picard. 2003. “ADELFE: A Methodology for Adaptive Multi-Agent Systems Engineering.” In Springer, Berlin, Heidelberg, 156–69.

Biernaskie, Jay M, and Stuart A West. 2015. “Cooperation, Clumping and the Evolution of Multicellularity.” *Proceedings. Biological sciences* 282(1813): 20151075.

Blythe, Richard A., and Thomas C. Scott-Phillips. 2014. “Simulating the Real Origins of Communication” ed. Stephanie Ann White. *PLoS ONE* 9(11): e113636.

Bonjean, N. et al. 2014. “ADELFE 2.0.” In *Handbook on Agent-Oriented Design Processes*, Berlin, Heidelberg: Springer Berlin Heidelberg, 19–63.

Börlin, Christoph S, Verena Lang, Anne Hamacher-Brady, and Nathan R Brady. 2014. “Agent-Based Modeling of Autophagy Reveals Emergent Regulatory Behavior of Spatio-Temporal Autophagy

Dynamics." *Cell Communication and Signaling* 12(1): 56.

Boulianne, Laurier, Sevin Al Assaad, Michel Dumontier, and Warren J Gross. 2008. "GridCell: A Stochastic Particle-Based Biological System Simulator." *BMC Systems Biology* 2(1): 66.

Bousquet, François, Christophe Le Page, and Jean-Pierre Müller. 2002. "Modélisation et Simulation Multi-Agent." In *Deuxiemes Assises Du GDRI3*,.

Brini, Emiliano et al. 2017. "How Water's Properties Are Encoded in Its Molecular Structure and Energies." *Chemical reviews* 117(19): 12385–414.

Bryan, Nathan S, Ka Bian, and Ferid Murad. 2009. "Discovery of the Nitric Oxide Signaling Pathway and Targets for Drug Development." *Frontiers in bioscience (Landmark edition)* 14: 1–18.

Caccavale, Justin et al. 2017. "A Simple and Accurate Rule-Based Modeling Framework for Simulation of Autocrine/paracrine Stimulation of Glioblastoma Cell Motility and Proliferation by L1CAM in 2-D Culture." *BMC Systems Biology* 11(1): 124.

Calabrese, Marco, Alberto Amato, Vincenzo Di Lecce, and Vincenzo Piuri. 2010. "Hierarchical-Granularity Holonic Modelling." *Journal of Ambient Intelligence and Humanized Computing* 1(3): 199–209.

Camps, Valérie, Marie-Pierre Gleizes, and Pierre Glize. 1998. "Une Théorie Des Phénomènes Globaux Fondée Sur Des Interactions Locales." In *Journées Francophones IAD SMA, Nancy,* , Hermès, 207–20.

Carrera, Javier, and Markus W Covert. 2015. "Why Build Whole-Cell Models?" *Trends in cell biology* 25(12): 719–22.

Chatterjee, Manash S, Jeremy E Purvis, Lawrence F Brass, and Scott L Diamond. 2010. "Pairwise Agonist Scanning Predicts Cellular Signaling Responses to Combinatorial Stimuli." *Nature biotechnology* 28(7): 727–32.

Cickovski, T.M. et al. 2005. "A Framework for Three-Dimensional Simulation of Morphogenesis." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2(4): 273–88.

Claessen, Dennis et al. 2014. "Bacterial Solutions to Multicellularity: A Tale of Biofilms, Filaments and Fruiting Bodies." *Nature Reviews Microbiology* 12(2): 115–24.

Clairambault, Jean. 2013. "Partial Differential Equation (PDE), Models." In *Encyclopedia of Systems Biology*, New York, NY: Springer New York, 1635–1635.

Clayton, Philip, and Paul Davies. 2006. *The Re-Emergence of Emergence : The Emergentist Hypothesis from Science to Religion*. Oxford University Press.

Cooper, Geoffrey M., and Robert E. Hausman. 2009. *The Cell : A Molecular Approach*. 5th ed. Washington D.C. ;Sunderland Mass.: ASM Press.

Crank, John. 1975. *The Mathematics of Diffusion*. Clarendon Press.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London : John Murray, 1859.

Deutsch, Andreas. 2016. *Cellular Automaton Modeling of Biological Pattern Formation : Characterization, Examples, and ... Analysis.* Birkhauser.

Diggle, Stephen P., Ashleigh S. Griffin, Genevieve S. Campbell, and Stuart A. West. 2007. "Cooperation and Conflict in Quorum-Sensing Bacterial Populations." *Nature* 450(7168): 411–14.

Ditlevsen, Susanne, and Adeline Samson. 2013. "Introduction to Stochastic Models in Biology." In Springer, Berlin, Heidelberg, 3–35.

Dobrescu, R, and V I Purcarea. 2011. "Emergence, Self-Organization and Morphogenesis in Biological Structures." *Journal of medicine and life* 4(1): 82–90.

Driscoll, William W, and Michael Travisano. 2017. “Synergistic Cooperation Promotes Multicellular Performance and Unicellular Free-Rider Persistence.” *Nature Communications* 8: 15707.

Duran-Nebreda, Salva et al. 2016. “Emergence of Proto-Organisms from Bistable Stochastic Differentiation and Adhesion.” *Journal of the Royal Society, Interface* 13(117).

Evans, James P, Eric M Meslin, Theresa M Marteau, and Timothy Caulfield. 2011. “Genomics. Deflating the Genomic Bubble.” *Science (New York, N.Y.)* 331(6019): 861–62.

Feig, Michael et al. 2015. “Complete Atomistic Model of a Bacterial Cytoplasm for Integrating Physics, Biochemistry, and Systems Biology.” *Journal of molecular graphics & modelling* 58: 1–9.

Feng, Cheng, and Jane Hillston. 2014. “PALOMA: A Process Algebra for Located Markovian Agents.” In Springer, Cham, 265–80.

Ferber, Jacques. 1999. *Multi-Agent Systems : An Introduction to Distributed Artificial Intelligence*. Addison-Wesley.

Floreano, Dario, Sara Mitri, Stéphane Magnenat, and Laurent Keller. 2007. “Evolutionary Conditions for the Emergence of Communication in Robots.” *Current Biology* 17(6): 514–19.

Flores, Enrique, and Antonia Herrero. 2010. “Compartmentalized Function through Cell Differentiation in Filamentous Cyanobacteria.” *Nature Reviews Microbiology* 8(1): 39–50.

Forrest, S. 1993. “Genetic Algorithms: Principles of Natural Selection Applied to Computation.” *Science (New York, N.Y.)* 261(5123): 872–78.

Friedmann, Elfriede. 2015. “PDE/ODE Modeling and Simulation to Determine the Role of Diffusion in Long-Term and -Range Cellular Signaling.” *BMC Biophysics* 8(1): 10.

Frigg, R., and S. Hartmann. 2012. *Models in Science*. Stanford University.

Frisch, U., B. Hasslacher, and Y. Pomeau. 1986. “Lattice-Gas Automata for the Navier-Stokes Equation.” *Physical Review Letters* 56(14): 1505–8.

Gardner, Timothy S., Charles R. Cantor, and James J. Collins. 2000. “Construction of a Genetic Toggle Switch in Escherichia Coli.” *Nature* 403(6767): 339–42.

Gaudet, Pascale et al. 2017. “The neXtProt Knowledgebase on Human Proteins: 2017 Update.” *Nucleic Acids Research* 45(D1): D177–82.

Georgé, Jean-Pierre, Bruce Edmonds, and Pierre Glize. 2004. “Making Self-Organising Adaptive Multiagent Systems Work.” In *Methodologies and Software Engineering for Agent Systems*, Boston: Kluwer Academic Publishers, 321–40.

Georgé, Jean-Pierre, Marie-Pierre Gleizes, and Pierre Glize. 2003. “Conception de Systèmes Adaptatifs À Fonctionnalité Émergente : La Théorie Amas.” *Revue d’intelligence artificielle* 17(4): 591–626.

Ghaffarizadeh, Ahmadreza et al. 2018. “PhysiCell: An Open Source Physics-Based Cell Simulator for 3-D Multicellular Systems” ed. Timothée Poisot. *PLOS Computational Biology* 14(2): e1005991.

Ghaffarizadeh, Ahmadreza, Samuel H. Friedman, and Paul Macklin. 2016. “BioFVM: An Efficient, Parallelized Diffusive Transport Solver for 3-D Biological Simulations.” *Bioinformatics* 32(8): 1256–58.

Gisler, Monika, Didier Sornette, and Ryan Woodard. 2010. “Exuberant Innovation: The Human Genome Project.”

Glize, Pierre. 2001. “L’adaptation Des Systèmes À Fonctionnalité Émergente Par Auto-Organisation Coopérative.” Habilitation à diriger des recherches, Université Paul Sabatier.

Goldberg, Arthur P. et al. 2018. “Emerging Whole-Cell Modeling Principles and Methods.” *Current Opinion in Biotechnology* 51: 97–102.

Goldstein, Jeffrey. 1999. “Emergence as a Construct: History and Issues.” *Emergence* 1(1): 49–72.

Graner, François, and James A. Glazier. 1992. “Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model.” *Physical Review Letters* 69(13): 2013–16.

Gratie, Diana-Elena, Bogdan Iancu, and Ion Petre. 2013. “ODE Analysis of Biological Systems.” In Springer, Berlin, Heidelberg, 29–62.

Griffiths, Anthony JF et al. 2000. “Bacterial Conjugation.”

Grosberg, Richard K, and Richard R Strathmann. 2007. “The Evolution of Multicellularity: A Minor Major Transition?” *Annu. Rev. Ecol. Evol. Syst* 38: 621–54.

Grouchy, Paul, Gabriele M. T. D’Eleuterio, Morten H. Christiansen, and Hod Lipson. 2016. “On The Evolutionary Origin of Symbolic Communication.” *Scientific Reports* 6(1): 34615.

Guespin-Michel, Janine. 2017. “La Révolution Du Complexe.” http://www.revolutionducomplexe.fr/.

H. Koschwanez, John, Kevin R. Foster, and Andrew W. Murray. 2011. “Sucrose Utilization in Budding Yeast as a Model for the Origin of Undifferentiated Multicellularity” ed. Laurent Keller. *PLoS Biology* 9(8): e1001122.

Handly, L Naomi, Anna Pilko, and Roy Wollman. 2015. “Paracrine Communication Maximizes Cellular Response Fidelity in Wound Signaling.” *eLife* 4: e09652.

Hardy, Simon, and Pierre N Robillard. 2004. “Modeling and Simulation of Molecular Biology Systems Using Petri Nets: Modeling Goals of Various Approaches.” *Journal of bioinformatics and computational biology* 2(4): 595–613.

Hartmann, Boris M et al. 2014. “Combinatorial Cytokine Code Generates Anti-Viral State in Dendritic Cells.” *Frontiers in immunology* 5: 73.

Heiner, Monika, David Gilbert, and Robin Donaldson. 2008. “Petri Nets for Systems and Synthetic Biology.” In *Formal Methods for Computational Systems Biology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 215–64.

Hoehme, Stefan, and Dirk Drasdo. 2010. “A Cell-Based Simulation Software for Multi-Cellular Systems.” *Bioinformatics* 26(20): 2641–42.

Hoops, S. et al. 2006. “COPASI--a COmplex PAthway SImulator.” *Bioinformatics* 22(24): 3067–74.

Hsueh, Robert C et al. 2009. “Deciphering Signaling Outcomes from a System of Complex Networks.” *Science signaling* 2(71): ra22.

Hwang, Minki, Marc Garbey, Scott A Berceli, and Roger Tran-Son-Tay. 2009. “Rule-Based Simulation of Multi-Cellular Biological Systems-A Review of Modeling Techniques.” *Cellular and molecular bioengineering* 2(3): 285–94.

Ispolatov, Iaroslav, Martin Ackermann, and Michael Doebeli. 2012. “Division of Labour and the Evolution of Multicellularity.” *Proceedings. Biological sciences* 279(1734): 1768–76.

Jacobeen, Shane et al. 2017. “Cellular Packing, Mechanical Stress and the Evolution of Multicellularity.” *Nature Physics*.

Jaynes, Jessica et al. 2013. “Application of Fractional Factorial Designs to Study Drug Combinations.” *Statistics in medicine* 32(2): 307–18.

Ji, Zhiwei et al. 2017. “Mathematical and Computational Modeling in Complex Biological Systems.” *BioMed Research International* 2017: 1–16.

Justice, Sheryl S., David A. Hunstad, Lynette Cegelski, and Scott J. Hultgren. 2008. “Morphological Plasticity as a Bacterial Survival Strategy.” *Nature Reviews Microbiology* 6(2): 162–68.

Kainkaryam, Raghunandan M, and Peter J Woolf. 2009. “Pooling in High-Throughput Drug Screening.” *Current opinion in drug discovery & development* 12(3): 339–50.

Kalenka, Susanne, and Nicholas R. Jennings. 1999. “Socially Responsible Decision Making by Autonomous Agents.” In *Cognition, Agency and Rationality*, Dordrecht: Springer Netherlands,

135–49.

Kamerlin, Shina C. L., Pankaz K. Sharma, Ram B. Prasad, and Arieh Warshel. 2013. “Why Nature Really Chose Phosphate.” *Quarterly Reviews of Biophysics* 46(1): 1–132.

Karr, Jonathan R. et al. 2012. “A Whole-Cell Computational Model Predicts Phenotype from Genotype.” *Cell* 150(2): 389–401.

Kauffman, S. 1969. “Homeostasis and Differentiation in Random Genetic Control Networks.” *Nature* 224(5215): 177–78.

Kim, Hong Pyo, Stefan W. Ryter, and Augustine M.K. Choi. 2006. “CO AS A CELLULAR SIGNALING MOLECULE.” *Annual Review of Pharmacology and Toxicology* 46(1): 411–49.

Kimura, Hideo. 2015. “Signaling Molecules: Hydrogen Sulfide and Polysulfide.” *Antioxidants & Redox Signaling* 22(5): 362–76.

Kirk, D. L. 2003. “Seeking the Ultimate and Proximate Causes of Volvox Multicellularity and Cellular Differentiation.” *Integrative and Comparative Biology* 43(2): 247–53.

Kirk, David L. 2005. “A Twelve-Step Program for Evolving Multicellularity and a Division of Labor.” *BioEssays* 27(3): 299–310.

Krenning, G. et al. 2013. “Combined Implantation of CD34 $^{+}$ and CD14 $^{+}$ Cells Increases Neovascularization through Amplified Paracrine Signalling.” *Journal of Tissue Engineering and Regenerative Medicine* 7(2): 118–28.

Krenning, Guido, Marja J.A. van Luyn, and Martin C. Harmsen. 2009. “Endothelial Progenitor Cell-Based Neovascularization: Implications for Therapy.” *Trends in Molecular Medicine* 15(4): 180–89.

Kugler, Hillel, Antti Larjo, and David Harel. 2010. “Biocharts: A Visual Formalism for Complex Biological Systems.” *Journal of the Royal Society, Interface* 7(48): 1015–24.

Kurve, Aditya, Khashayar Kotobi, and George Kesidis. 2013. “An Agent-Based Framework for Performance Modeling of an Optimistic Parallel Discrete Event Simulator.” *Complex Adaptive Systems Modeling* 1(1): 12.

Kwon, Hyuk Min et al. 2014. “Multiple Paracrine Factors Secreted by Mesenchymal Stem Cells Contribute to Angiogenesis.” *Vascular Pharmacology* 63(1): 19–28.

Laio, Alessandro, and Michele Parrinello. 2002. “Escaping Free-Energy Minima.” *Proceedings of the National Academy of Sciences of the United States of America* 99(20): 12562–66.

Libby, Eric, Peter L Conlin, Ben Kerr, and William C Ratcliff. 2016. “Stabilizing Multicellularity through Ratcheting.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371(1701).

Libby, Eric, and William C Ratcliff. 2014. “Evolution. Ratcheting the Evolution of Multicellularity.” *Science (New York, N.Y.)* 346(6208): 426–27.

Lipkow, Karen, Steven S Andrews, and Dennis Bray. 2005. “Simulated Diffusion of Phosphorylated CheY through the Cytoplasm of Escherichia Coli.” *Journal of bacteriology* 187(1): 45–53.

Lipson, Hod. 2007. “Evolutionary Robotics: Emergence of Communication.” *Current Biology* 17(9): R330–32.

Lodish, Harvey et al. 2000. “Overview of Extracellular Signaling.” In *Molecular Cell Biology (4th Ed.)*, New York: W. H. Freeman and Co.

Lodish, Harvey, Arnold Berk, and S Lawrence Zipursky. 2000. “The Life Cycle of Cells.” In *Molecular Cell Biology (4th Ed.)*, New York: W. H. Freeman and Co.

Lodish, Harvey F. 2013. *Molecular Cell Biology*. (7th ed.). New York: W.H. Freeman and Co.

Lukeš, Julius et al. 2011. “How a Neutral Evolutionary Ratchet Can Build Cellular Complexity.” *IUBMB*

Life 63(7): 528–37.

Lyons, Nicholas A, and Roberto Kolter. 2015. “On the Evolution of Bacterial Multicellularity.” *Current opinion in microbiology* 24: 21–28.

Mano, Jean-Pierre, Christine Bourjot, Gabriel Lopardo, and Pierre Glize. 2006. “Bio-Inspired Mechanisms for Artificial Self-Organised Systems.” *Informatica* 30(1): 55--62.

Di Marzo Serugendo, Giovanna, Marie-Pierre Gleizes, and Anthony Karageorgos. 2011. “Self-Organising Systems.” In Springer, Berlin, Heidelberg, 7–32.

McClellan, James Scott et al. 2015. “Reprogramming of Primary Human Philadelphia Chromosome-Positive B Cell Acute Lymphoblastic Leukemia Cells into Nonleukemic Macrophages.” *Proceedings of the National Academy of Sciences of the United States of America* 112(13): 4074–79.

McPartland, Michelle, Stefano Nolfi, and Hussein A. Abbass. 2005. “Emergence of Communication in Competitive Multi-Agent Systems.” In *Proceedings of the 2005 Conference on Genetic and Evolutionary Computation - GECCO ’05*, New York, New York, USA: ACM Press, 51.

Meier-Schellersheim, Martin, Iain D C Fraser, and Frederick Klauschen. 2009. “Multiscale Modeling for Biologists.” *Wiley interdisciplinary reviews. Systems biology and medicine* 1(1): 4–14.

Milholland, Brandon et al. 2017. “Differences between Germline and Somatic Mutation Rates in Humans and Mice.” *Nature Communications* 8: 15183.

Miller, Melissa B., and Bonnie L. Bassler. 2001. “Quorum Sensing in Bacteria.” *Annual Review of Microbiology* 55(1): 165–99.

Mina, Petros, Krasimira Tsaneva-Atanasova, and Mario di Bernardo. 2016. “Entrainment and Control of Bacterial Populations: An *in Silico* Study over a Spatially Extended Agent Based Model.” *ACS Synthetic Biology* 5(7): 639–53.

Mirams, Gary R. et al. 2013. “Chaste: An Open Source C++ Library for Computational Physiology and Biology” ed. Andreas Prlic. *PLoS Computational Biology* 9(3): e1002970.

Mitri, Sara, Dario Floreano, and Laurent Keller. 2011. “Relatedness Influences Signal Reliability in Evolving Robots.” *Proceedings. Biological sciences* 278(1704): 378–83.

Moe-Behrens, Gerd Hg. 2013. “The Biological Microprocessor, or How to Build a Computer with Biological Parts.” *Computational and structural biotechnology journal* 7: e201304003.

Moraru, I I et al. 2008. “Virtual Cell Modelling and Simulation Software Environment.” *IET systems biology* 2(5): 352–62.

Natarajan, Madhusudan et al. 2006. “A Global Analysis of Cross-Talk in a Mammalian Cellular Signalling Network.” *Nature Cell Biology* 8(6): 571–80.

Niklas, K. J. 2014. “The Evolutionary-Developmental Origins of Multicellularity.” *American Journal of Botany* 101(1): 6–25.

Nikolić, Đurica, Corrado Priami, and Roberto Zunino. 2012. “A Rule-Based and Imperative Language for Biochemical Modeling and Simulation.” In Springer, Berlin, Heidelberg, 16–32.

Noël, Victor, and Franco Zambonelli. 2015. “Methodological Guidelines for Engineering Self-Organization and Emergence.” In Springer, Cham, 355–78.

Picard, Gauthier. 2004. “Méthodologie de Développement de Systèmes Multi-Agents Adaptatifs et Conception de Logiciels À Fonctionnalité Émergente.”

Picard, Gauthier, and Marie-Pierre Gleizes. 2004. “The ADELFE Methodology.” In *Methodologies and Software Engineering for Agent Systems*, Boston: Kluwer Academic Publishers, 157–75.

Qiao, Wenlian et al. 2014. “Intercellular Network Structure and Regulatory Motifs in the Human Hematopoietic System.” *Molecular systems biology* 10(7): 741.

Quinn, Matt. 2001. “Evolving Communication without Dedicated Communication Channels.” In Springer, Berlin, Heidelberg, 357–66.

Quinn, Matt, Lincoln Smith, Giles Mayley, and Phil Husbands. 2003. “Evolving Controllers for a Homogeneous System of Physical Robots: Structured Cooperation with Minimal Sensors.” *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 361(1811): 2321–43.

Raychaudhuri, Subhadip. 2013. “Kinetic Monte Carlo Simulation in Biophysics and Systems Biology.” In *Theory and Applications of Monte Carlo Simulations*, InTech.

Redou, Pascal et al. 2007. “Formal Validation of Asynchronous Interaction-Agents Algorithms for Reaction-Diffusion Problems.” In *21st International Workshop on Principles of Advanced and Distributed Simulation*, IEEE, 93–104.

Rejniak, Katarzyna A. 2007. “An Immersed Boundary Framework for Modelling the Growth of Individual Cells: An Application to the Early Tumour Development.” *Journal of theoretical biology* 247(1): 186–204.

Rivière, J., P. Ballet, and V. Rodin. 2016. “NetBioDyn, a Smart Agent-Based Software to Intuitively Model and Simulate Complex Biological Systems.” In *CCS 2016, the 2016 Conference on Complex Systems*,.

Rueffler, Claus, Joachim Hermisson, and Günter P Wagner. 2011. “Evolution of Functional Specialization and Division of Labor.” *Pnas* 109(6): E326–35.

Ruiz-Trillo, Iñaki et al. 2007. “The Origins of Multicellularity: A Multi-Taxon Genome Initiative.” *Trends in Genetics* 23(3): 113–18.

Russell, Stuart J. (Stuart Jonathan), and Peter. Norvig. 1995. *Artificial Intelligence : A Modern Approach*. Prentice Hall.

Scholes, Natalie S., and Mark Isalan. 2017. “A Three-Step Framework for Programming Pattern Formation.” *Current Opinion in Chemical Biology* 40: 1–7.

Schrier, Sarah B., Abby S. Hill, Deborah Plana, and Douglas A. Lauffenburger. 2016. “Synergistic Communication between CD4+ T Cells and Monocytes Impacts the Cytokine Environment.” *Scientific Reports* 6(1): 34942.

Shao, Hongwei, Ying He, King C. P. Li, and Xiaobo Zhou. 2013. “A System Mathematical Model of a Cell–cell Communication Network in Amyotrophic Lateral Sclerosis.” *Molecular BioSystems* 9(3): 398.

Shimomura, Koji. 1998. “A Dynamic Equilibrium Model of Durable Goods Monopoly.” *Journal of Economic Behavior & Organization* 33(3–4): 507–20.

Sims, Christopher E., and Nancy L. Allbritton. 2007. “Analysis of Single Mammalian Cells on-Chip.” *Lab on a Chip* 7(4): 423.

Starruß, Jörn, Walter de Back, Lutz Brusch, and Andreas Deutsch. 2014. “Morpheus: A User-Friendly Modeling Environment for Multiscale and Multicellular Systems Biology.” *Bioinformatics* 30(9): 1331–32.

Sutterlin, T. et al. 2013. “Bridging the Scales: Semantic Integration of Quantitative SBML in Graphical Multi-Cellular Models and Simulations with EPISIM and COPASI.” *Bioinformatics* 29(2): 223–29.

Szabó, András, and Roeland M H Merks. 2013. “Cellular Potts Modeling of Tumor Growth, Tumor Invasion, and Tumor Evolution.” *Frontiers in oncology* 3: 87.

Tanaka, Simon, David Sichau, and Dagmar Iber. 2015. “LBIBCell: A Cell-Based Simulation Environment for Morphogenetic Problems.” *Bioinformatics* 31(14): 2340–47.

Tao Peng et al. 2014. “Modeling Cell–Cell Interactions in Regulating Multiple Myeloma Initiating Cell Fate.” *IEEE Journal of Biomedical and Health Informatics* 18(2): 484–91.

Thurley, Kevin, Daniel Gerecht, Elfriede Friedmann, and Thomas Höfer. 2015. “Three-Dimensional Gradients of Cytokine Signaling between T Cells” ed. Rustom Antia. *PLOS Computational Biology* 11(4): e1004206.

Thurley, Kevin, Lani F. Wu, and Steven J. Altschuler. 2018. “Modeling Cell-to-Cell Communication Networks Using Response-Time Distributions.” *Cell Systems* 0(0).

Tomar, Namrata, Olivia Choudhury, Ankush Chakrabarty, and Rajat K. De. 2013. “An Integrated Pathway System Modeling of Saccharomyces Cerevisiae HOG Pathway: A Petri Net Based Approach.” *Molecular Biology Reports* 40(2): 1103–25.

Topin, Xavier et al. 1999. “Theories and Experiments on Emergent Behaviour : From Natural to Artificial Systems and Back.” In *European Conference on Cognitive Science, Siena*,.

Tuljapurkar, S D, and J S Semura. 1977. “Dynamic Equilibrium under Periodic Perturbations in Simple Ecosystem Models.” *Journal of Theoretical Biology* 66(2): 327–43.

Ulieru, Mihaela, and Rene Doursat. 2011. “Emergent Engineering: A Radical Paradigm Shift.” *International Journal of Autonomous and Adaptive Communications Systems* 4(1): 39.

Vivas, J, D Garzón-Alvarado, and M Cerrolaza. 2015. “Modeling Cell Adhesion and Proliferation: A Cellular-Automata Based Approach.” *Advanced modeling and simulation in engineering sciences* 2: 32.

Voss-Böhme, Anja. 2012. “Multi-Scale Modeling in Morphogenesis: A Critical Analysis of the Cellular Potts Model.” *PloS one* 7(9): e42852.

Walpole, J. et al. 2015. “Agent-Based Model of Angiogenesis Simulates Capillary Sprout Initiation in Multicellular Networks.” *Integrative Biology* 7(9): 987–97.

Weiss, Gerhard. 1999. *Multiagent Systems : A Modern Approach to Distributed Artificial Intelligence*. MIT Press.

Weyns, Danny et al. 2005. “Environments for Multiagent Systems State-of-the-Art and Research Challenges.” In Springer, Berlin, Heidelberg, 1–47.

Wilensky, U. 1999. “NetLogo.”

Wilkinson, D. J. 2006. “Bayesian Methods in Bioinformatics and Computational Systems Biology.” *Briefings in Bioinformatics* 8(2): 109–16.

Winsberg, Eric. 2015. “Computer Simulations in Science.” In *The Stanford Encyclopedia of Philosophy*, ed. Edward N Zalta. Metaphysics Research Lab, Stanford University.

De Wolf, Tom, and Tom Holvoet. 2005. “Emergence Versus Self-Organisation: Different Concepts but Promising When Combined.” In Springer, Berlin, Heidelberg, 1–15.

Zhang, Zheren, Dennis Claessen, and Daniel E Rozen. 2016. “Understanding Microbial Divisions of Labor.” *Frontiers in microbiology* 7: 2070.

Personal Bibliography (by date)

Maignan S., Bernon C. and Glize P. 2018. "Speeding up the Search of a Global Dynamic Equilibrium from a Local Cooperative Decision." In Proceedings of the 10th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART: 143-150.

Carry, Jean-Christophe et al. 2015. "SAR156497, an Exquisitely Selective Inhibitor of Aurora Kinases." *Journal of Medicinal Chemistry* 58(1): 362–75.

Liang, Guyan et al. 2012. "Structure-Based Library Design and the Discovery of a Potent and Selective Mast Cell β-Tryptase Inhibitor as an Oral Therapeutic Agent." *Bioorganic & Medicinal Chemistry Letters* 22(2): 1049–54.

Lesuisse, Dominique et al. 2011. "Discovery of the First Non-ATP Competitive IGF-1R Kinase Inhibitors: Advantages in Comparison with Competitive Inhibitors." *Bioorganic & Medicinal Chemistry Letters* 21(8): 2224–28.

Nemecek, Conception et al. 2010. "Design of Potent IGF1-R Inhibitors Related to Bis-Azaindoles." *Chemical Biology & Drug Design* 76(2): 100–106.

Lesuisse, Dominique et al. 2010. "Rational Design of Potent GSK3β Inhibitors with Selectivity for Cdk1 and Cdk2." *Bioorganic & Medicinal Chemistry Letters* 20(6): 1985–89.

Hopkins, Corey R. et al. 2005. "Design, Synthesis, and Biological Activity of Potent and Selective Inhibitors of Mast Cell Tryptase." *Bioorganic & Medicinal Chemistry Letters* 15(11): 2734–37.

Levell, Julian et al. 2005. "Structure Based Design of 4-(3-Aminomethylphenyl)piperidinyl-1-Amides: Novel, Potent, Selective, and Orally Bioavailable Inhibitors of βII Tryptase." *Bioorganic & Medicinal Chemistry* 13(8): 2859–72.

Hopkins, Corey et al. 2004. "Novel Pyrazinone Inhibitors of Mast Cell Tryptase: Synthesis and SAR Evaluation." *Bioorganic & Medicinal Chemistry Letters* 14(19): 4819–23.

Maignan, Sébastien et al. 2003. "Molecular Structures of Human Factor Xa Complexed with Ketopiperazine Inhibitors: Preference for a Neutral Group in the S1 Pocket." *Journal of Medicinal Chemistry* 46(5): 685–90.

Choi-Sledeski, Yong Mi et al. 2003. "Discovery of an Orally Efficacious Inhibitor of Coagulation Factor Xa Which Incorporates a Neutral P_1 Ligand." *Journal of Medicinal Chemistry* 46(5): 681–84.

Guertin, Kevin R et al. 2002. "Optimization of the Beta-Aminoester Class of Factor Xa Inhibitors. Part 2: Identification of FXV673 as a Potent and Selective Inhibitor with Excellent In Vivo Anticoagulant Activity." *Bioorganic & medicinal chemistry letters* 12(12): 1671–74.

Maignan, S, and V Mikol. 2001. "The Use of 3D Structural Data in the Design of Specific Factor Xa Inhibitors." *Current topics in medicinal chemistry* 1(2): 161–74.

Maignan, S et al. 2000. "Crystal Structures of Human Factor Xa Complexed with Potent Inhibitors." *Journal of medicinal chemistry* 43(17): 3226–32.

Choi-Sledeski, Y M et al. 1999. "Sulfonamidopyrrolidinone Factor Xa Inhibitors: Potency and Selectivity Enhancements via P-1 and P-4 Optimization." *Journal of medicinal chemistry* 42(18): 3572–87.

Ewing, William R. et al. 1999. "Design and Structure−Activity Relationships of Potent and Selective Inhibitors of Blood Coagulation Factor Xa." *Journal of Medicinal Chemistry* 42(18): 3557–71.

Maignan, Sébastien et al. 1998. "Crystal Structures of the Catalytic Domain of HIV-1 Integrase Free and Complexed with Its Metal Cofactor: High Level of Similarity of the Active Site with Other Viral Integrases 1 1Edited by R. Huber." *Journal of Molecular Biology* 282(2): 359–68.

Erskine, P.T. et al. 1997. "Crystallization of 5-Aminolaevulinic Acid Dehydratase from *Escherichia Coli* and *Saccharomyces Cerevisiae* and Preliminary X-Ray Characterization of the Crystals." *Protein Science* 6(8): 1774–76.

Vaney, M. C., S. Maignan, M. Riès-Kautt, and A. Ducruix. 1996. “High-Resolution Structure (1.33 Å) of a HEW Lysozyme Tetragonal Crystal Grown in the APCF Apparatus. Data and Structural Comparison with a Crystal Grown under Microgravity from SpaceHab-01 Mission.” *Acta Crystallographica Section D Biological Crystallography* 52(3): 505–17.

Chardin, P, D Cussac, S Maignan, and A Ducruix. 1995. “The Grb2 Adaptor.” *FEBS letters* 369(1): 47–51.

Maignan, S et al. 1995. “Crystal Structure of the Mammalian Grb2 Adaptor.” *Science (New York, N.Y.)* 268(5208): 291–93.