BIOLOGICAL ORGANISATION AND VIABILITY:
A THEORY OF AUTOPOIESIS IN CELLULAR AND EUCLIDEAN AUTOMATA

Thomas M. Gaul

Submitted to the faculty of the Cognitive Science Program
in partial fulfillment of the requirements for
Departmental Honors
in the degree of Bachelor of Science
in the Cognitive Science Program
Indiana University
April 2025

Accepted by the Cognitive Science Faculty, Indiana University, in partial fulfillment of the requirements for conferral of departmental Honors in the Cognitive Science Program

Thesis Committee

Randall D. Beer, Ph.D.

Eduardo J. Izquierdo, Ph.D.

Luis H. Favela, Ph.D.

April 9, 2025

Acknowledgements

My path leading up to this thesis is not something I could have anticipated at the beginning of my undergraduate career, less than four years ago. But I would not have achieved what I have were it not for the many people whose patience and generosity I have depended on. Unfortunately, I can only thank a few here.

I would first like to thank Gabe Severino for taking the time to listen to a freshman's naïve speculations about consciousness, and for introducing me to enaction, Wittgenstein, and dynamical approaches to cognitive science, all of which deeply invigorated my excitement to pursue my curiosities seriously and scientifically. I would also like to thank Ann Barwich for encouraging my intellectual development throughout innumerable after-class conversations.

I thank Eduardo Izquierdo for advising me on my first research project, and for introducing me to the Computational Neuroethology Lab. I thank my advisor, Randy Beer, for his invaluable guidance and patience in my journey to becoming researcher. I thank Connor McShaffrey for the countless conservations, and for his detailed feedback on earlier drafts of this thesis.

I thank the members of my committee: Randy Beer, Eduardo Izquierdo, and Louie Favela. I thank the members of the Computational Neuroethology Lab: Connor McShaffrey, Gabe Severino, Eden Forbes, Lindsay Stolting, Denizhan Pak, Haily Merritt, and Zach Laborde. I thank the many people who make up the Cognitive Science Program at Indiana University.

Finally, I thank my parents for their enduring support of all my endeavours.

Thomas M. Gaul

Biological Organisation and Viability: A Theory of Autopoiesis in Cellular and Euclidean Automata

Increasingly, biologists are emphasising the cell as the most fundamental unit of life, and importantly, a view of the cell as an emergent, dynamic object irreducible to its molecular composition – "greater than the sum of its parts." But what makes a cell itself *living*? Most answers to this question are unsatisfactory: lists of properties or behaviours that always have exceptions. Moreover, there is a lack of consensus that does not appear to be near resolution.

The theory of *autopoiesis* offers a potential solution here, providing an operational definition of the *organisation* of a cell, independent of what molecules compose it. Autopoietic theory allows us to answer questions about when a cell dies, and how it can behave. However, for it to realise this utility, rigorous theories need to be developed for specific models of emergent individuals. Beer (2020b) has previously used the *Game of Life* cellular automaton as one such model, and developed a mathematical theory of autopoiesis that can predict the behaviour and death of an emergent individual (Beer, McShaffrey, & Gaul, 2024).

This thesis extends and reformulates this work to a broader class of cellular automata, called *Larger than Life* (LtL) (Evans, 1996). I then further extend the LtL theory to a Euclidean automata, called *RealLife* (Pivato, 2007), and provide evidence of the convergence of autopoiesis in LtL to RealLife. I also introduce a method for deriving automata rules from the organisation of an individual.

Finally, I discuss the implications autopoietic theory has for our understanding of life and cognition, including a reevaluation of the relation between autopoiesis and the enactive approach to cognition (Di Paolo, 2005; Varela et al., 2017).

Randall D. Beer, Ph.D.

Eduardo J. Izquierdo, Ph.D.

Luis H. Favela, Ph.D.

Contents

List of Figures

Chapter 1

Introduction

Biology encompasses a vast range of phenomena, from the molecular details of bacterial metabolism, to the intelligent behaviour of corvids; from the evolution of language, to the reciprocal symbiosis of coral and algae. But despite the great diversity of life apparent all around us, there is just as clearly something that unifies these phenomena – something that makes them distinctly *biological*. The history of biology, especially since the nineteenth century, can be read as the history of attempts at unifying our understanding of all those things we call living (Morange, 2021; Varela, 1979). But how far have we come? With the dominance of molecular biology in the twentieth century, many came to see the problem as solved: living systems were carriers of a "genetic code," and all of their characteristics could be successfully explained in terms of simple physics and chemistry, and the origin of these characteristics by natural selection on genes (Dawkins, 1976; Judson, 2013; Mayr, 1961; Monod, 1971; Morange, 2005, 2010). On the other hand, there were some who saw the question of delineating life from non-life as simply beyond the purview of scientific concerns (Foucault, 1975; Jacob, 1982; Morange, 2010).

Most biologists now recognise both of these positions as problematic. The reductive molecular view of life is unable to account for the emergence of complex phenotypes that cannot be attributed to the simple action of genes (Kirschner et al., 2000; Woese, 2004). Instead, *cells* are commonly seen as the most basic unit of life (Mukherjee, 2022). Our understanding of cells themselves has also changed: they are complex, self-organising dynamic patterns undergoing constant material turnover of their components (Hodgins, 2009; Kirschner et al., 2000; Margulis & Sagan, 2000; Wilmsen & Kost, 2025). Some even see life as being distinguished by purposeful agency (Ball, 2023). It is also clear that certain problems in biology cannot be solved without a clearer notion of what exactly it means for something to be living, such as the origins of life on Earth (Ruiz-Mirazo & Moreno, 2004) and life on extraterrestrial planets (Colón-Santos et al., 2024).

However, despite the reinvigoration of attempts to answer the question “what is life?”, there is hardly any consensus on what that answer is, or even what it looks like (Colón-Santos et al., 2024; Cornish-Bowden, 2006, 2011; Wilmsen & Kost, 2025). But there are still some general characteristics that are often seen in proposed definitions of life: organisms have complex metabolism, they are spatially delimited, they adapt to stimuli, they reproduce with variation (or, they have the capacity for Darwinian evolution), they are self-organising systems, they exhibit growth, they are homeostatic, etc. (Colón-Santos et al., 2024; Morange, 2005; Mukherjee, 2022; Nurse, 2021; Pross, 2016; Rasmussen et al., 2022; Wilmsen & Kost, 2025). All of these, each on their own or in any combination, is unsatisfactory. The most basic problem with most of them is that they describe what life *does*, but not what it *is*. For instance, it is certainly true that organisms adapt to stimuli in their environment, but this does not tell us how to demarcate the organism from that environment. Similarly, the mere fact of reproduction does not say *what* is being reproduced, and evolution does not say what kinds of populations are being evolved (if natural selection can be applied to universes (Smolin, 2004), does that make them living?). There are also clear exceptions to some of these characteristics: mules – and, for that matter, most elderly humans – cannot reproduce, but are they not still living things? On the other hand, the less problematic characteristics are still quite vague, e.g., what does it mean for a system to be "self-organising"?

Another class of definitions, more common in origins of life and astrobiology research (Colón-Santos et al., 2024), is to define life by the biomolecules we observe on Earth, such as proteins, nucleic acids, etc. But such definitions fail to sufficiently take into account the material turnover mentioned above. If the particular molecules that compose an organism at any given time can be replaced, and thus do not determine whether that organism is alive, why should classes of molecules be seen any different? Put simply, it is the *dynamic form* of an organism that matters, not what instantiates that form.

The insufficiency of all these definitions can be made even clearer by asking a simple question: *when does an organism die*? Clearly, a molecular definition is insufficient here since all would agree that a random mixture of biomolecules does not constitute a living system. On the other hand, the list of properties given above are either too general, or not relevant. There are non-biological systems that are self-organising, that reproduce, etc., and properties like being a product of evolution have little bearing on how we demarcate whether a system is dead *here and now*.

What we need, then, is a more precise operational definition of what makes something living – a notion of the emergence and identity of biological individuals as persistent dynamic forms irreducible to the components that realise it at any given time. It is thus the task of theoretical biology to provide a rigorous theory of *biological identity* – an answer to the question: what makes a living system *living*?[1] Importantly, such a theory must, as a consequence of the irreducibility to its components, contain an *identity-realisation duality*: the organism is logically distinct from the molecules that compose it. This perspective reframes many biological questions, making explicit the distinction between the space of *possibilities* and the contingency of particular realisations. This allows us to ask new questions about living systems:

[1] I want to be clear that I do not interpret this as a matter of defining life *per se*, but of providing an explicit account of (some of) *those systems that we call living*. That is, the theory I seek is not meant to police language of what we should or should not call living, but to provide an alternative and more rigorous way of talking that helps to answer specific questions and clarifies our conceptual frameworks.

- What is it made of? $\implies$ What are its possible realisations?
- What does it do? $\implies$ What can it do?
- When does it die? $\implies$ When is its identity no longer realised?

To be sure, not every question in biology requires such a definitions, nor is it necessarily bad that most current proposals are unsatisfactory. It is useful to have guiding intuitions about how we should conceptualise biological problems, even if such intuitions are at this point imprecise. Moreover, a pluralistic approach to understanding life may prove more immediately practical, as is the case in astrobiology (Colón-Santos et al., 2024). But if we accept the basic tenet that the whole range of diverse biological phenomena forms a unified intelligible class, distinct from other domains of science, then it is imperative that biology, as a whole, have a theoretical framework that defines its subject matter precisely. What I propose is that a theory of biological identity fulfills this demand.

1.0.1. Theories of Biological Identity

There is a rich history in theoretical biology of attempts at formulating a proper theory of biological identity. This is sometimes traced all the way back to Kant's introduction of the term "self-organisation" (Kant, 1790/2000; Weber & Varela, 2002). But it was not until later in the twentieth century that serious scientific efforts were dedicated to the question of biological identity, at least in its current manifestation (Cornish-Bowden and Cárdenas (2020) gives a brief history and overview of some of these theories). Some of the most prevalent theories still referenced today include (M,R) systems (Rosen, 1991), hypercycles (Eigen & Schuster, 1979), chemotons (Gánti, 2003), autopoiesis (Maturana & Varela, 1980), autocatalytic sets (Kauffman, 1986), and closure of constraints (Montévil & Mossio, 2015).

These all certainly deserve further development in their own right, especially as they each emphasise different aspects of what makes living systems unique. But one has to make a choice on which to develop, and so I have chosen *autopoiesis* (Maturana & Varela, 1980; Varela, 1979). I do this because autopoiesis has the potential to be an especially powerful theory with implications far beyond that of biology. Foremost among these is an intimate connection between identity and cognition, and thus a grounding of cognitive science in biology. It also emphasises *viability* to a greater extent than other theories: the capacity of a system to survive, and what it means for it to die (Bourgine & Varela, 1992; Di Paolo, 2005, 2009; Varela, 1979). This therefore puts autopoiesis in position to be a very useful theory to address scientific problems concerning behaviour, viability, and the relation these have to the self-maintaining processes that constitute the individual as a distinct biological identity.

1.1. Epistemology

Autopoietic theory relies on a few basic epistemological concepts concerning how observers (or more properly, observer-communities; Varela, 1979) make descriptions. Most simply, a description consists of unities and their properties/relations. Distinct classes of unities and relations constitute distinct *descriptive domains*, or *domains of description* – basically, ways of describing things. For instance, we can consider a particle-level quantum mechanical description of a chemical reaction, and a stoichiometric one. In the quantum mechanical one, the basic unities are particles with quantum states, related to one another by their position in various fields. In the stoichiometric description, the basic unities are species of molecules, with concentrations, rates of change, etc. These descriptive domains are, strictly speaking, incommensurate with one another, in the sense that the phenomena of one are not proper to the other – individual particles do not have concentrations.

An observer can also mediate between descriptive domains by mapping unities in one to collections of unities in another, giving rise to *composite unities*. Thus, an observer can consider different descriptions of the same system, [2] and associate phenomena in one with

[2] Of course, what grounds "sameness" here will depend on a further descriptive domain, ultimately grounded in the pragmatics of embodied observers making descriptions. A fuller discussion of this would involve unravelling the whole epistemological framework here, but for my purposes, it will be sufficient to simply say that two descriptions are of the same system when manipulating that system involves manipulating the same place

the other – that is, explaining one phenomenon in terms of the other, on the grounds of the map between their respective descriptive domains.

With this, we can understand the most important concept in autopoietic theory, besides that of autopoiesis itself: the distinction between *organisation* and *structure*. The structure of a system is simply all of the physical relations between, and properties of, its components through time. The organisation of the system is any subset or abstraction of those relations. Thus, we say that a structure *realises* a particular organisation when a subset of its relations are precisely those of the organisation. This implies that a single structure can realise many different organisations; intuitively, this is analogous to describing the same dog, for example, as *this* dog, a dog, a mammal, an animal, etc.

When the unities and relations generated by such a process of abstraction in turn generate a qualitatively distinct domain of description, we say this new domain is *emergent* with respect to the original from which it was abstracted. Thus, this is a relatively weak sense of emergence in that it is necessarily relative to an observer, and any interaction between distinct descriptive domains is necessarily mediated by an observer.

In all, then, this epistemological framework allows us to characterise systems first and foremost by their organisation, such that they can be realised in space, yet remain distinct from their structure at any given moment. Moreover, the observer plays a central role in thinking about this dichotomy, and especially in mediating between different domains of description. The central task of theoretical biology is then to determine the *organisation* of living systems, and to explain how the constraints implied by its realisation in turn constrain the phenomena proper to different domains of description, whether those be molecular, behavioural, ecological, or evolutionary. This is precisely what autopoietic theory aims to do.

in physical space.

1.2. Autopoiesis

Maturana and Varela proposed autopoiesis as the defining organization of living systems (Maturana & Varela, 1980; Varela, 1979; Varela et al., 1974). The original definition is quite verbose:

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously generate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network. (Maturana & Varela, 1980, pp. 78–79).

More intuitively, we can think of an autopoietic system as a network of processes, analogous to chemical reactions, with two properties:

1. (**Self-Production**) The network continuously regenerates itself through the components it produces,
2. (**Self-Distinction**) The network produces and maintains a spatial boundary that it depends on.[3]

Thus, the term auto- (self) poiesis (production) (from the Greek: αυτός/self, ποιειν/to produce). The paradigmatic example of an autopoietic system is the single cell. Its autopoietic organisation consists of all the reactions (processes) that produce the components of the cell (DNA, proteins, lipids, sugars, etc.) and the reactions that produce the reactants of these processes. Among the components that this network produces are phospholipids that constitute the plasma membrane distinguishing the cell from its environment and

[3] Technically, an autopoietic system need not exist in physical space, according to the original definition. However, given its primary application to livings systems, physical space is almost always assumed. In any case, Maturana (2002) made clear his view that autopoiesis could only exist in the molecular domain.

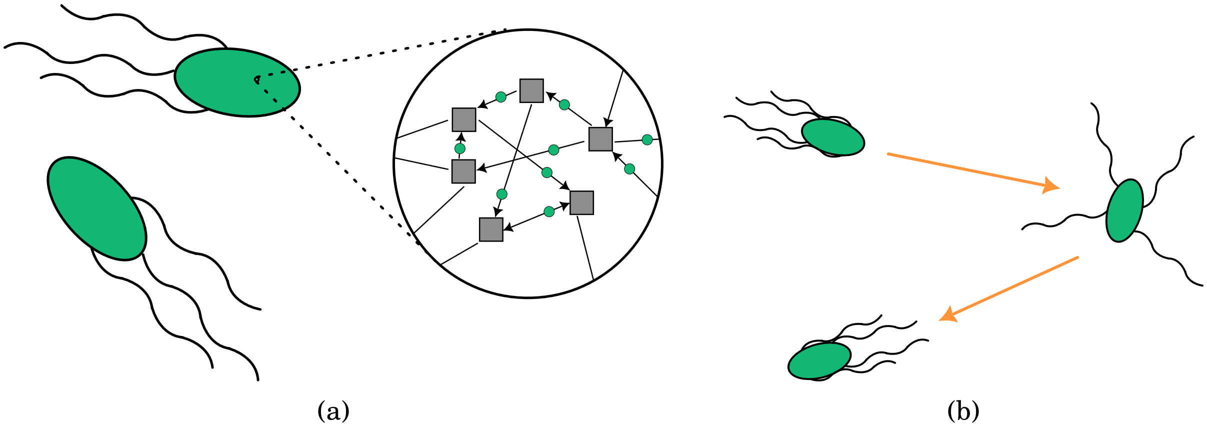


Figure 1.1: Schematic illustration of the autopoiesis of *E. coli*. (a) An illustration of part of the autopoietic organisation of *E. coli*. Each gray box is a process (a chemical reaction). Arrows between processes processes represent the production by one process of a reactant (green) in the other. (b) An illustration of the 'run and tumble' behaviour of *E. coli*.

preventing the disintegration of the network by diffusion. To make this even more concrete, consider the schematic illustration of the bacterium *Escherichia coli* in Figure 1.1a. Here, the gray boxes represent reactions/processes, and arrows represent the exchange of components whereby one process produces a molecule (green circle) that then becomes a reactant in another process. Every molecule that constitutes the cell structure at any given instance is understood to be realising a relation between processes. Thus, the existence of DNA molecules, proteins complexes, the plasma membrane, etc., are implicit in the network as intermediate enabling relations. The power of autopoiesis lies in how this implicit specification allows us to talk both about living systems as independent from their structural realisations, and in terms of their *possible* realisations.

An important interpretative point about this definition is the precariousness of autopoietic systems (Beer & Di Paolo, 2023). In its original form, the definition allows trivial cases in which none of the processes of the network really depend on any other processes for their continued operation. In contrast, it is clear that any chemical reaction within a cell is dependent on other reactions, most obviously so with respect to the enzymes needed for most reactions to proceed at the necessary rates.

Notably absent from the definition of autopoiesis is any notion of reproduction or evolution. This is because, as mentioned above with respect to evolution, these concepts are logically dependent on the identity of the reproducing systems. That is, there is *something* reproduced, and that something is precisely the autopoietic organisation. Another way to say this is that autopoiesis is a systemic characterisation of living systems, not a list of properties in relation to a context outside the system itself.

As noted above, autopoiesis has many far-reaching implications for how we think of various scientific problems, including cognition, viability, evolution, the origins of life, extraterrestrial life, phenomenology, and epistemology, just to name a few. I will choose to focus on cognition and viability, as these constitute, with autopoiesis itself, the essential core of autopoietic theory upon which all the other implications rest.

1.2.1. Cognition

Because an autopoietic system distinguishes itself as a unity in space, we can also describe it as a single behaving system. A central claim of autopoietic theory is that the organisation of a system determines the space of its possible behaviours. This is because the organisation specifies what structural transitions are permissible with respect to itself. That is, at any given instant, some structural transitions will destroy the organisation, while other will preserve it. But these internal structural transitions can just as well be seen from a different perspective as the behaviour of the system. To illustrate this, consider again *E. coli* (Figure 1.1b). Behaviourally, an *E. coli* bacterium will swim toward higher concentrations of nutrients (“run”) and will randomly change directions (“tumble”) until a positive gradient is found (Sourjik & Wingreen, 2012). These behaviours are regulated by protein modifications and a system of signalling proteins that ultimately influences the flagellar motors. From an autopoietic perspective, these protein interactions are different concatenations of processes in the organisation. Thus, the organisation, by describing all the enabling relations between processes, implies the possibility of both behaviours as different realisations of the network.

In this way, the organisation establishes a perspective on the world by differentiating environmental conditions based on what permissible structural transitions they induce. *E. coli* distinguishes between lower and higher concentrations of nutrient because it behaves differently in each circumstance. Put another way, the realisation of the *E. coli* organisation gives rise to different behaviours depending on the environment in which that realisation is situated, and this differentiation is determined by the organisation. The whole space of survivable perturbations forms a system's *domain of interaction*.

This combined implication of behaviour and perspective is what justifies the claim that autopoietic systems are cognitive systems (Maturana & Varela, 1980) — the domain of domain of interaction then becomes the *cognitive domain*. That is, cognition is understood here to be the behaviour and perspective of an autopoietic system. This is different from how cognition is generally understood in cognitive science (Boden, 2000), or even by others in the autopoietic literature (Di Paolo, 2009; Thompson, 2007). For instance, Di Paolo (2005) understands cognition to require that the system not only establish a perspective, but also a norm or goal with respect to which it acts. However, it is worth bearing in mind that we need not call all autopoietic systems cognitive, as Maturana and Varela are not claiming that all autopoietic systems exhibit the full richness of cognitive behaviour observed in animals. Thus, if one wishes to use the word "cognition" more strictly, this is fine, so long as they recognise the relation between behaviour and perspective implied by autopoiesis and the continuity of this relation from cells to humans (Maturana & Varela, 1987). For this thesis, I will use the word cognition the same way as in the original literature, i.e., all autopoietic systems and cognitive systems.

The structure-organisation distinction also makes clear the co-determination of behaviour by by both the organism and its environment: the organisation does not determine the behaviour of the individual, but only the possibilities in any given moment. For behaviour to emerge, the environment must select paths through the permitted structural transitions, through a process called *structural coupling* (Figure 1.2). Thus, the organisation of *E. coli* does not determine whether it will run or tumble, until it is embedded in a particular environment. Structural coupling is also what give rise to the historical dimen-

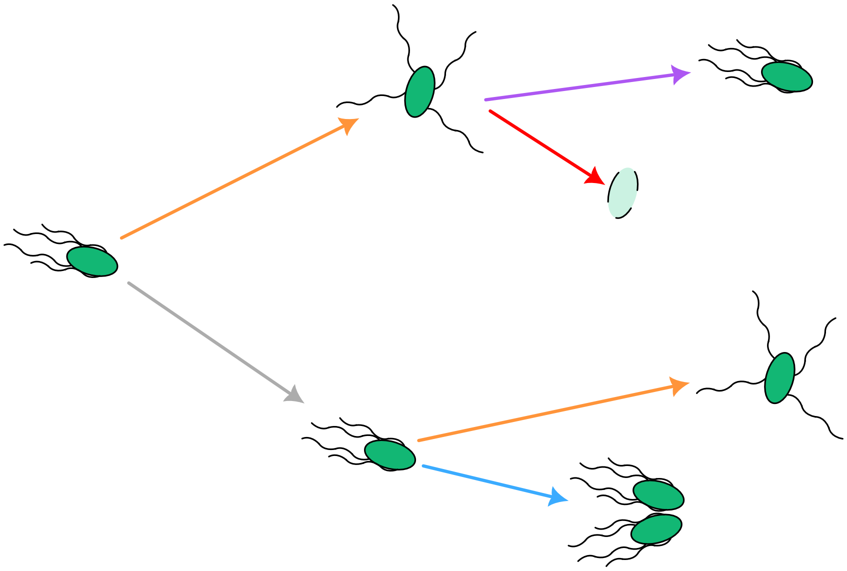


Figure 1.2: A schematic illustration of structural coupling in *E. coli*. Each arrow represents a different transition between behaviours. Gray is the continuation of running. Orange is the transition for running to tumbling. Blue is the transition from running to biolfilm formation or quorum sensing. Red is the transition from tumbling to death. Purple is the transition from tumbling to running.

sion of living systems, for the paths of structural transitions that an autopoietic system takes is dependent upon sequences of environmental conditions external to the organisation (though the space of *possible* paths is again determined by the organisation).

1.2.2. Viability

By much of the same reasoning as for cognition, the autopoiesis of an individual will also delimit the environmental conditions that destroy its organisation, and thus kill it. Similarly, organisation determines the viable structural configurations of an individual. Again, using *E. coli* as an example, the organisation would determine the space of possible antibiotics, nutrient concentrations, etc., that could kill the cell, and the particular structure the cell must be in under those conditions, while also strictly delimiting the precise timing of when the cell is considered to be dead (i.e., when the organisation is destroyed).

Viability in autopoietic theory is often understood as it is formulated by Ashby (1960).[4]

[4] Aubin (1991) is also a relevant source of ideas on viability and in the context of control theory, but his work is referenced much less in the literature on autopoiesis (but see Bourgine and Varela, 1992).

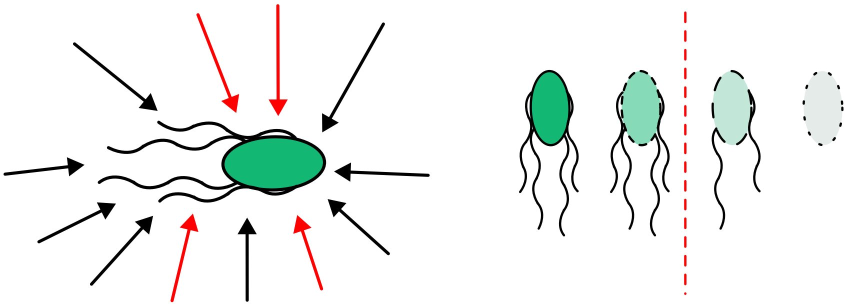


Figure 1.3: Schematic illustration of viability in *E. coli*. On the left, the black arrows represented environments that do not destroy the organisation, while red arrows do. The organisation of *E. coli* determines what environments belong to which group. On the right is a sequence of *E. coli* with increasingly broken membranes and lower internal concentration. The organisation determines the exact point at which the bacterium is considered dead (the red line).

Ashby employed the notion of *essential variables* as the variables of the system whose values determine the whether the system is viable: if any essential variable exceeds some range of values, then the system is no longer viable (it is non-functioning, or dead). These ranges of values give rise to a *viability region* where all of the essential variables are within their acceptable range. The boundary of this region is where any essential variable is at the maximum or minimum of its range. Applied to *E. coli*, the essential variables might be the concentration of glucose or lactose within the cell, the internal osmotic pressure, the pH within the cell, etc. If any of these variables exceed certain limits, the bacterium will die. This is often how viability is formulated in computational biology, independent of Ashby (e.g., Ghaffarizadeh et al., 2018).

A prevalent concept in current autopoietic theory related to viability is *adaptivity*, introduced by Di Paolo (2005). It is meant to be an extension of autopoiesis in order to account for agency and *sense-making*: a valenced perspective on the world as meaningful for the system (Thompson, 2007; Varela et al., 2017). Adaptivity is defined as the capacity of a system to regulate itself and its relation to the environment so as to avoid approaching the boundary of viability too closely. This is claimed to be a necessary and sufficient condition for an autopoietic system to establish for itself intrinsic norms that it behaves with respect to – *E. coli* swims up a a nutrient gradient *because* it is good for the cell (Di Paolo, 2005,

p. 437). This is also taken to imply an intentionality whereby a individual is directed toward its own future (Di Paolo, 2005, p. 443). Adaptivity, then, is a bridge from autopoiesis as originally defined to the enactive conception of cognition as the bringing forth of a meaningful world, and its intimate relation to phenomenology (Di Paolo, 2005, 2009; Di Paolo et al., 2018; Thompson, 2007; Varela et al., 2017).

There are of course still controversies concerning how precisely to interpret the definition of autopoiesis. Must the boundary necessarily be a physical boundary? Can the system extend beyond its physical boundary? Can the organisation of a system change over time? Are cells the only instance of an autopoietic system? Must every component in the network be produced by the network? Is autopoiesis both necessary *and* sufficient for life? Does autopoiesis really imply "cognition" (Boden, 2000; Di Paolo, 2009; Fleischaker, 1988; Luisi, 2003; McMullin, 1999; Virgo et al., 2011)?

Much of this discussion, however, has been carried out without reference to any precise definitions of the basic concepts of autopoiesis. Thus, depending on how one implicitly understands autopoiesis, they may rightly come to a completely different conclusion from another, simply because they have a different understanding of, say, what "process" means. To alleviate this confusion, I feel that it would be best to use a model where my interpretation of autopoiesis can be made explicit and precise. In doing so, the above problems can be made more concrete and tractable.

1.3. Modelling and Formalisation

Using a modelling approach allows us not only to address theoretical questions in concrete terms (Beer, 2020d), but also offers a way of developing the theoretical and technical tools needed for a mature mathematical theory with predictive power (Beer, Barwich, & Severino, 2024). Thus, if autopoiesis is to fully realise its potential, modelling should figure as a key part in its development.

Fortunately, there is a rich history of modelling autopoiesis, and biological organisation more generally. For my purposes here, it will be useful to split these models into two classes: spatial and non-spatial. Non-spatial models have the advantage of being easier to analyse and to achieve rigorous results with. Moreover, such models are prevalent across systems biology, including whole-cell models (Alon, 2019; Karr et al., 2012; Sun et al., 2021).

However, with respect to questions concerning emergent biological identity, non-spatial models are unsatisfactory. For instance, organisation is almost always non-emergent in these models, but instead is specified by the observer. Most often, this is in the form a system of ordinary differential equations (ODEs) describing the organisation as an autocatalytic network of reactions (Barandiaran & Egbert, 2014; Bourgine & Stewart, 2004; McShaffrey & Beer, 2023). More obviously, these models lack a physical spatial boundary that the system produces to distinguish itself from the environment. In some cases, there is no physical boundary at all (Fontana & Buss, 1994), or the boundary is assumed as an implicit condition for the autocatalytic network to exist (Barandiaran & Egbert, 2014; McShaffrey & Beer, 2023), or else it is imposed on the system by the observer (Bourgine & Stewart, 2004). The predetermination of the organisation by the observer also means that the organisation cannot, strictly speaking, disintegrate. Thus, to investigate viability in these models, the constraint delimiting the space of viable states is imposed on the system by the observer, often as bounds on the essential variables (Ashby, 1960).

Spatial models, on the other hand, more readily exhibit a diverse space of phenomena, including the emergence of biological identity. However, what one gains in complexity, one loses in the ability to attain rigorous mathematical results, often having to compensate with many computational simulations of very select situations (Agmon et al., 2016; Egbert & Di Paolo, 2009; McMullin, 2004; Ono & Ikegami, 2000; Varela et al., 1974). More importantly, analyses of these models have never properly *defined* the organisation of a system, despite such being properly emergent. That is, the network of processes that constitutes an individual is never specified beyond the underlying chemistry that itself need not give rise to an emergent identity. In other words, there is no *theory* of biological organisation developed for these models.

Despite these shortcomings, it is clear that a spatial model is ultimately necessary to properly develop a theory of autopoiesis that captures all of the relevant features (i.e., the emergence of organisation and the production of a physical boundary). The main challenge, then, is to determine what a mathematical theory of autopoiesis should look like in these models. To start, we can list some desiderata that such a theory should satisfy (at least with respect to the core problems discussed above: cognition and viability). It should precisely define what a *process* is. It should explicitly define and represent the organisation of an individual; we should have a concrete mathematical object describing the autopoietic organisation of a system independent of any realisation of it. From this organisation, we should be able to derive the complete space of behaviour possible for the system. Finally, we should be able to derive the complete space of environments that destroy the organisation, and the complete space of environments that do not.

There is one exception to the general trend of spatial models that does achieve this. Beer has used the Game of Life (GoL) cellular automata (Adamatzky, 2010; Berlekamp et al., 1982) as a simple universe in which a formal theory of autopoiesis has been developed (Beer, 2004, 2014, 2015, 2020a, 2020b, 2020c; Beer, McShaffrey, & Gaul, 2024). The GoL universe is a discrete two-dimensional lattice of cells that can be either on or off (Figure 1.4). Every cell is updated simultaneously by taking the sum of the on-cells in the immediate neighbourhood of the cell, and using the following rule:

- If the cell is off and the sum of on-cells in the neighbourhood is equal to 3, the cell will turn on; otherwise, it will stay off.

- If the cell is on and the sum of on-cells in the neighbourhood is greater than 2 and less than 3, the cell will stay on; otherwise, the cell will turn off.

This simple rule gives rise to a surprising diversity of persistent emergent patterns, such as *blocks*, *blinkers*, and *gliders* (Figure 1.4), which fall into the more general classes

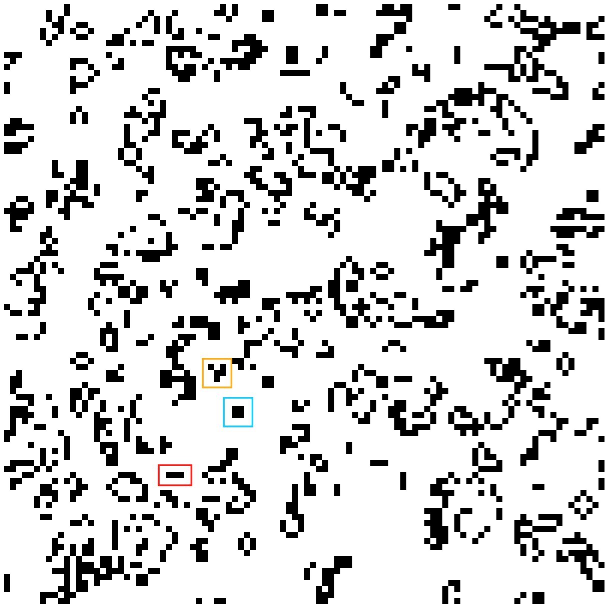


Figure 1.4: A snapshot of a GoL universe simulation. Black cells are on-cells, white cells are off. In orange is boxed a *glider*. in Blue us boxed a *block*. In red is boxed a *blinker*.

of *still-lifes*, *oscillators*, and *spaceships*, respectively.

In order to use GoL as a useful tool for developing autopoietic theory, Beer (2015) interpreted this update rule as a physics, which in turn gives rise to a chemistry. Specifically, a cell and all the cells in its neighbourhood can be interpreted as reactants that, through the update rule, produce a product: the updated cell (Figure 1.5a). The on- and off-cells of GoL then become simple molecules, called *0-* and *1-components*. A *process* is simply defined by a specific neighbourhood of components and the product they produce (Figure 1.5a). Depending on the center-component and the product, a process can classified into one of four classes: 1-maintenance (1 → 1), 0-maintenance (0 → 0), destruction (1 → 0), and production (0 → 1).

This chemistry itself gives rise to the emergent individuals mentioned above. However, to interpret them in autopoietic terms, they must have some kind of boundary separating them as an individual from the environment. To resolve this, Beer (2004) interprets the 0-components immediately surrounding the 1-components of a pattern as its boundary. For example, the glider is characterised by a specific pattern of five 1-components (Fig-

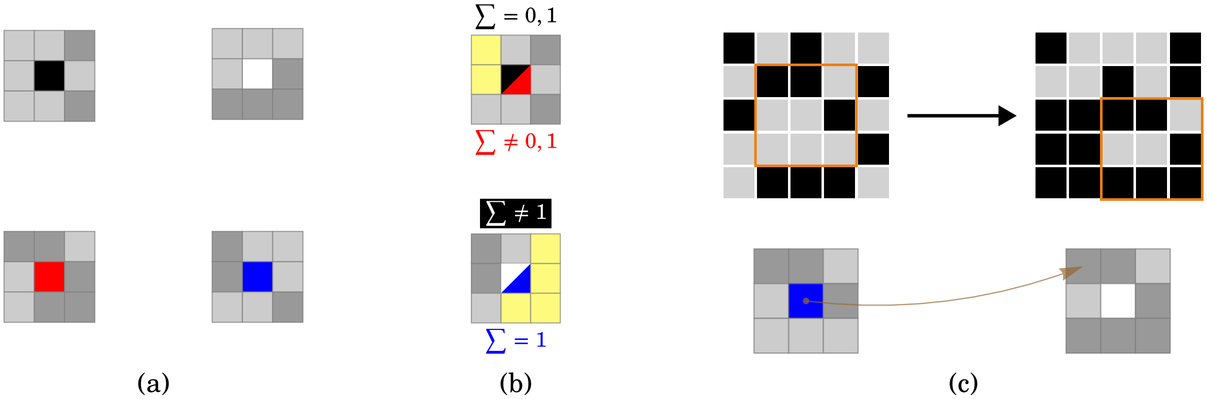


Figure 1.5: Beer's2020b process notation. (a) Four classes of processes. Light gray cells are 0-components, dark gray cells are 1-components. A black center-cell represents a *1-maintenance* process; white a *0-maintenance* process; red a *destruction* process; and blue a *production* process. (b) Partial processes. Yellow cells are unspecified components. A half-red half-black center-cell indicates a partial 1-maintenance/destruction process. If the sum of 1-components in the unspecified cells is 0 or 1, the process becomes 1-maintenance; otherwise, it becomes a destruction. A half-blue half-white center-cell indicates a partial 0-maintenance/production process. If the sum of 1-components in the unspecified cells is 1, the process becomes a production; otherwise, it becomes 0-maintenance. (c) Process dependency in GoL. The production process enables the 0-maintenance process because its product becomes a reactant at the next time step. The brown arrow indicates what role the product plays in the enabled process. (Figures (a) and (b) are adapted from Figure 2 of Beer (2020b, p. 204))

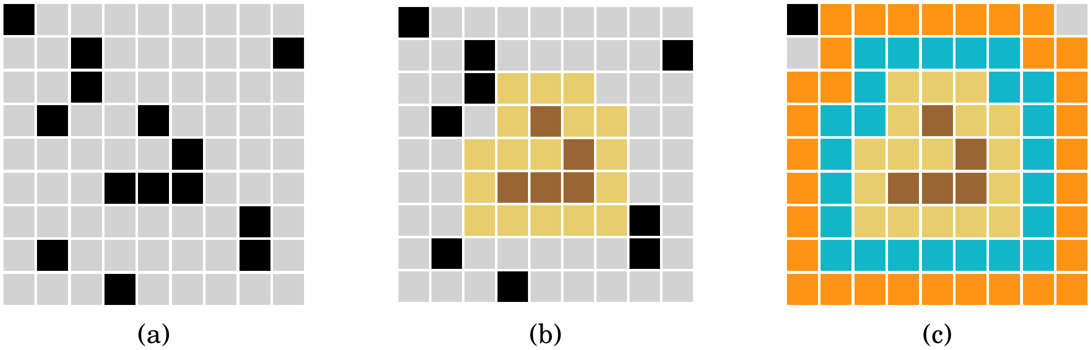
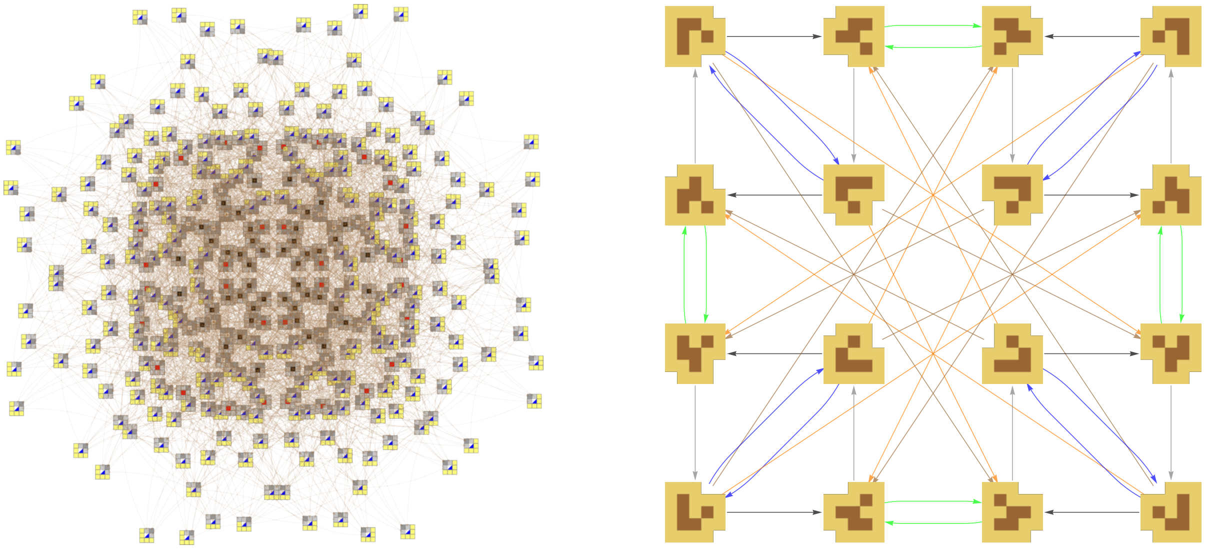


Figure 1.6: Glider configurations. Black cells are on, gray cells are off. (a) A glider configuration in a random environment. (b) A glider configuration with a boundary. Brown and beige cells are 1-components and 0-components, respectively, that belong to the organisation of the glider. (c) The (1,2)-environment of a glider. Blue cells are part of the glider's 1-environment; orange cells are part of the glider's 2-environment.

ure 1.6a). Taking all of the components in the neighbourhood of the pattern as part of the individual distinguishes it from the environment by isolating the internal 1-components from any external influence (Figure 1.6b). Thus, the boundary mediates all interaction with the surrounding milieu, while participating as reactants in the processes centered on the 1-components – the boundary both distinguishes the system as a unity and enables its continued maintenance. This separation also gives rise to a further structuring of the environment according to the speed at which influences can travel to the individual. Beer (2014) calls the components immediately surrounding the glider the *1-environment*, the components immediately surrounding these the *2-environment*, and so on (Figure 1.6c).

To formulate the autopoietic organisation of emergent individuals, Beer (2015) uses *process dependency networks* that describe all of the enabling relations between the processes centered on each component of the individual. One process enables another when the product of the first acts as a reactant in the other (Figure 1.5c). However, to account for the contingency of organisations embedded in environments, the notion of a *partial process* must be introduced (Beer, 2020b; Figure 1.5b). A partial process is one where at least one of the reactants are unspecified, indicated by a yellow cell. These processes can then be realised in more than one way, with each realisation resulting in a fully-determined process (Figure 1.5a). Moreover, depending on the number of 1-components among the unspecified



(a)

(b)

Figure 1.7: (a) The process dependency graph of a glider. (b) The interaction graph of a glider. The colored edges represent different transitions between glider configurations. Two edges have the same color of the transitions are symmetric. (These figures are reproductions of Figures 2 and 5c from Beer, McShaffrey, and Gaul (2024, p. 25, 28))

reactants of a process, the product may change (Figure 1.5b). In the context of an organisation, an unspecified component corresponds to the 1-environment of the individual.

Using this formalism, the autopoietic organisation of an emergent individual can be explicitly represented as a graph of processes and enabling relations (Figure 1.7a; Beer, 2020b). From this graph, one can derive (using an algorithm I implemented: Beer, McShaffrey, and Gaul, 2024) an *interaction graph* describing all the possible interactions an individual can engage in (Figure 1.7b; Beer, 2020b; Beer, McShaffrey, and Gaul, 2024), which can be verified by exhaustive computation (Beer, 2014). From an interaction graph, one can then partition the space of environments in which the organisation of an individual will exist, or will disintegrate (Figure 1.8). This constitutes the *intrinsic viability constraint* of an individual, in contrast to the *extrinsic* constraints in previous models, since the constraint is derived from considering how the organisation is embedded in space, instead of being imposed on the system by the observer as limits to essential variables.

The extrinsic-intrinsic distinction gives rise to a few interesting differences. First, the

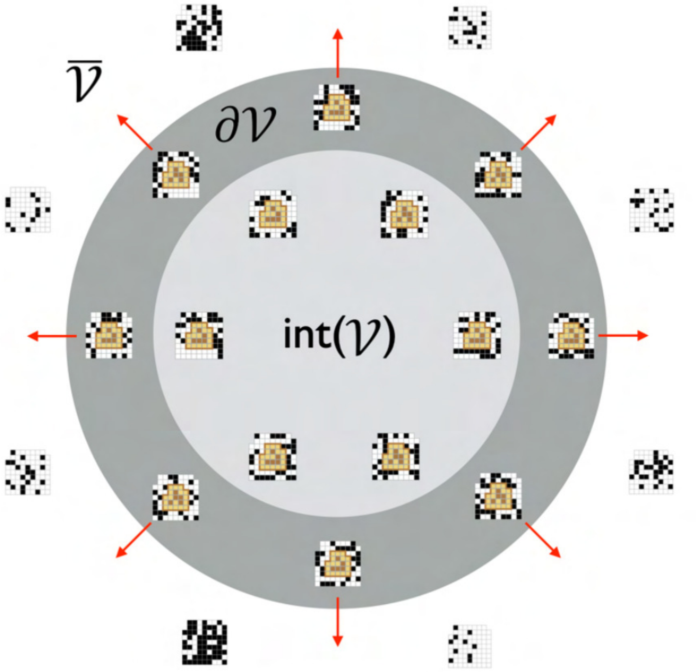


Figure 1.8: The intrinsic viability constraint of a glider. `int(`$\mathcal{V}$`)` contains glider (1,2)-environments that preserve the organisation (a glider will exist at the next time step). $\partial\mathcal{V}$ contains (1,2)-environments that destroy the organisation. $\overline{\mathcal{V}}$ contains configurations with no glider. (This figure is adapted from Figure 6 in Beer, McShaffrey, and Gaul (2024, p. 29))

information needed to determine the intrinsic constraint includes components in the 2-environment – components that are outside the immediate range of influence of the individual. Thus, in general, an autopoietic system cannot distinguish between environments in the boundary and in the interior of the constraint when those environments differ only by the 2-environment. Second, the essential variables of an individual may include environmental components, for the same reason. Third, the intrinsic constraint captures all possible behaviour. This contrasts with previous models where behavioural mechanisms were treated as something additional that did not modify the constraint (Barandiaran & Egbert, 2014). Fourth, there are no configurations in the boundary that return to the interior, since the configurations in the boundary, by definition, destroy the organisation.

Beer's work has shown that the main claims of autopoietic theory – the implication of a cognitive domain and the viability constraint from the organisation – hold true in the GoL universe. Moreover, the interpretation of autopoiesis formalised here involves essential physical boundaries and an inherent spatiality, both in the underlying chemistry and in the organisation it gives rise to. The formalisation of viability in GoL also challenges the concept of adaptivity and its dependence on "sensing the limits of viability." For instance, consider the intrinsic viability constraint of the glider (Figure 1.8). The glider cannot change its behaviour with respect to the viability constraint, since any such change would already be accounted for: the glider cannot alter a terminal trajectory. But even if we permitted such changes, the glider will not, in general, be able to determine whether it is on a terminal trajectory, since the information needed would be outside the 1-environment. Thus, adaptivity would appear to depend on the extrinsic formulation of viability, and a version of it that permits the modification of behaviour independent of the constraint.

***

A natural extension of Beer's work is to step toward a theory of autopoiesis in a universe more like our own, with continuous space, continuous time, thermodynamics, etc.

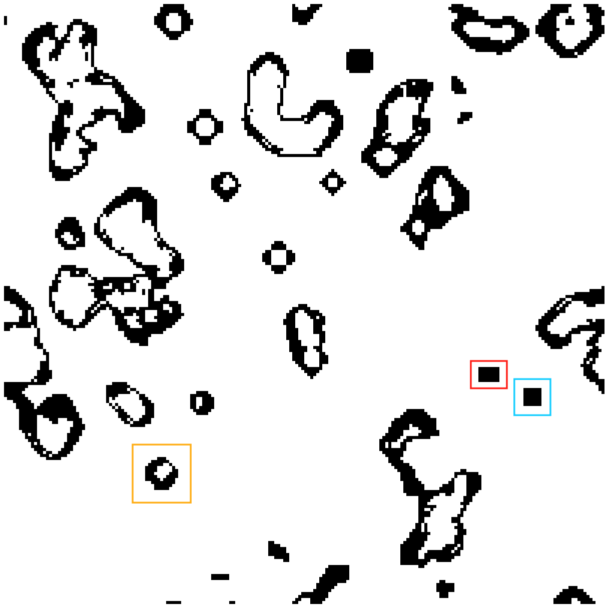


Figure 1.9: Snapshot of an LtL universe (under Bosco's rule: Evans, 2001). Black-cells are off, white-cells are on. In orange is boxed a *bug*. In red is boxed a *blinker*. In blue is boxed a *block*.

However, a more reasonable approach would be to choose one of these aspects and develop an extension of the theory there. Fortunately, there already exists rigorous results that connect GoL to a continuous space, discrete time universe. Specifically, *Larger than Life* (LtL) is a generalisation of GoL to arbitrarily large neighbourhoods that exhibits many similar emergent patterns (Figure 1.9; Evans, 1996, 2001, 2003, 2009) – GoL is an instance of LtL. Pivato (2007) has proven that there exists a continuum limit of LtL, called *RealLife*, as the neighbourhood size approaches infinity, by treating these larger neighbourhood rules as increasing the resolution of an approximation.

The programme I pursue then proceeds as follows. First, I will extend the interpretation of autopoiesis to LtL, before extending the theory and derivations. Then I will do the same for RealLife and establish a limiting connection between the discrete and continuous definitions of autopoiesis. I also present additional results, not present in the GoL work, for computing LtL/RealLife rules from an organisation. Finally, in the discussion chapter I interrogate what these results have to say about the challenge to adaptivity presented in Beer, McShaffrey, and Gaul (2024).

The next two chapters present the following results: (i) the proper object to describe autopoiesis in LtL is an edge-decorated digraph; (ii) some graphs in LtL demonstrate unique properties not observed in Beer's work on the glider; (iii) graphs in LtL converge to edge-decorated *graphons* in RealLife; (iv) interaction graphs can be derived algorithmically in both LtL and RealLife; (v) intrinsic viability constraints can be derived algorithmically in both LtL and RealLife; (vi) a set of rules that support the realisation of an organisation can be derived and, under certain conditions, this derivation is exhaustive; (vii) Beer's interpretation of autopoiesis holds in both LtL and RealLife; and (viii) the challenge to adaptivity holds in both LtL and RealLife.

In the final chapter, I discuss the significance these results have for our understanding of biological identity, cognition, and viability, and further pursue the implications the intrinsic formulation of viability has for adaptivity. I then sketch an alternative perspective to operationalise many of the explanatory targets of adaptivity. I end by considering future directions to this work and their potential applications.

Chapter 2

Autopoiesis in Larger than Life

This chapter develops a formal framework for analysing autopoiesis in families of Larger than Life (LtL) cellular automata (CA). First, I will define LtL and note some of the features that make it interesting and worth using as a model of emergent individuals. Next, I will broadly explain how persistent spatio-temporal patterns in LtL can be interpreted in autopoietic terms, before formulating this interpretation in a mathematical framework, illustrated with a few examples. The following sections extrapolate the consequences of this definition, first with respect to the cognitive domain of emergent individuals, and then with respect to the conditions of survival for those individuals. Rigorous definitions can be found in Appendix A.1.

2.1. Larger than Life

LtL is a family of two-dimensional outer-totalistic cellular automata (Evans, 2001, 2003) generalising the famous Game of Life (GoL) (Berlekamp et al., 1982). 'Outer totalistic' means that the rules governing how a given cell updates is determined by both the current state of the cell (outer) and the sum of the cells in its neighbourhood (totalistic). The neighbourhood of a cell consists of all the sites within the $(2\rho + 1) \times (2\rho + 1)$ box centred at the cell: the generalized Moore neighbourhood. All LtL rules share a basic form: if the cell is off and the sum over the neighbourhood falls within a contiguous *birth interval*, the cell will turn on; if the cell is on and the sum falls within a contiguous *survival interval*, the cell will stay on. If the sum does not fall into the interval, the cell will stay off or turn off, depending on whether the cell was off or on, respectively.

Formally, we denote the set of cell states $\mathscr{M} := \{0, 1\}$. The state of a universe is given by a function $\mathbf{u} : \mathbb{Z}^2 \to \mathscr{M}$, assigning to each site on a two-dimensional lattice a value in $\mathscr{M}$; I will notate the set of all these functions $\mathcal{U} \subset \mathscr{M}^{\mathbb{Z}^2}$. We can define the neighbourhood of a cell by $\mathscr{N} := \left\{ y : \|y\|_p \leq \rho, y \in \mathbb{Z}^2 \right\}$, where $\|\cdot\|_p$ is the p-norm and ρ is the neighbourhood radius; here, I will only consider neighbourhoods with $p = \infty$, i.e., the generalized Moore neighbourhood.

For the update rule, I will use the *average* value over the neighbourhood instead of the sum, which is equivalent to scaling the sum by a constant factor $1/|\mathscr{N}|$. Thus, we can define a convolution kernel κ to give us the *density* of on-cells in the neighbourhood of a point on the lattice x:

$$\kappa(y) := |\mathscr{N}|^{-1} \mathbb{1}_{\mathscr{N}}[y] \tag{2.1}$$

$$\kappa * \mathbf{u}(x) = |\mathscr{N}|^{-1} \sum_{y \in \mathscr{N}} \mathbf{u}(x - y). \tag{2.2}$$

We will often need to refer to the convolution $\kappa * \mathbf{u}$ (often calling it a *density function*) so we will notate it more compactly as $\Psi := \kappa * \mathbf{u}$.

Finally, an LtL CA is an operator $\xi : \mathcal{U} \to \mathcal{U}$, specified by a birth interval $[b_0, b_1] \subseteq [0, 1]$ and survival interval $[s_0, s_1] \subseteq [0, 1]$ such that

$$0 \leq s_0 \leq b_0 \leq b_1 \leq s_1 \leq 1, \tag{2.3}$$

defined by the following equation, where $\mathbb{1}_{\mathscr{A}}$ is the indicator function of the set $\mathscr{A}$:

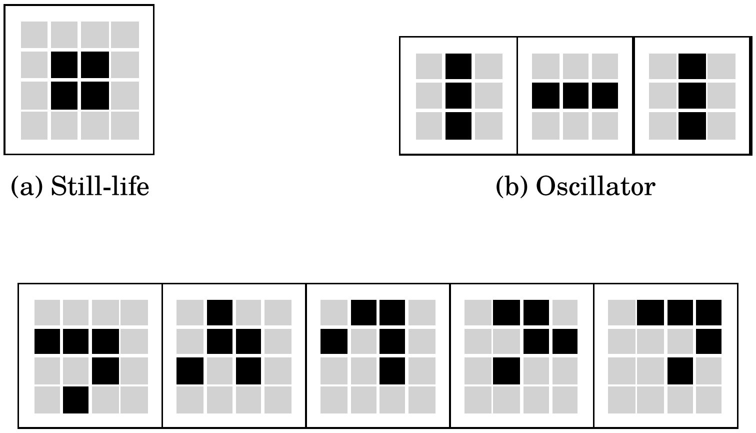
$$\xi[\mathbf{u}](x) := \mathbf{u}(x) \cdot \mathbb{1}_{[s_0,s_1]}[\Psi(x)] + (1 - \mathbf{u}(x)) \cdot \mathbb{1}_{[b_0,b_1]}[\Psi(x)]. \tag{2.4}$$

At each time step, the universe function updates according to ξ; thus, the time evolution of an LtL universe is determined by recursing ξ; I will notate this recursion by ξ^t for t time steps. LtL rules are usually specified by a 5-tuple $(\rho, b_0, b_1, s_0, s_1)$, suppressing dependence on the choice of p-norm. Hence, GoL is the rule $(1, \frac{3}{9}, \frac{3}{9}, \frac{3}{9}, \frac{4}{9})$. As one last bit of notation, we can express a spatial translation of a universe state as $(\sigma^v \circ \mathbf{u})(x) := \mathbf{u}(x-v)$ for all $x, v \in \mathbb{Z}^2$.

What makes LtL interesting for purposes of theoretical biology is the existence of emergent, precarious, recurrent, compact spatio-temporal patterns (Beer, 2004; Evans, 2001, 2003; Gaul, 2024). These patterns fall into three main classes: *still-lifes*, *oscillators*, and *bugs* (Evans, 2001). *Still-lifes* are patterns invariant with respect to ξ; they do not move or change. *Oscillators* are patterns that cycle through some finite number of configurations, where the smallest such cycle is the pattern's *period* (they are periodic with respect to ξ); still-lifes are a special case of oscillators with period 1. Finally, *bugs* are patterns that move by a given non-zero *translation vector* over their period – they are periodic mod spatial translation (note that these classifications apply to patterns surrounded complete with off-cells). We will be most interested in a particular sub-class of bugs – *bugs with stomachs* (see Figure 2.2c) – composed of a connected region of off-cells surrounded by a connected region of on-cells. These bugs are supported by a broader class of LtL rules, and their scaling with respect to ρ is more interesting (they do not tend to an infinitely thin configuration, for example; Evans, 2001).

More formally, a still-life is a universe state $\mathbf{u}$ such that $\xi[\mathbf{u}] = \mathbf{u}$; an oscillator is a state such that $\xi^{\bar{t}}[\mathbf{u}] = \mathbf{u}$ for some finite $\bar{t} \geq 1$; and a bug is a state such that $\xi^{\bar{t}}[\mathbf{u}] = \sigma^d \circ \mathbf{u}$ for some finite $\bar{t} \geq 1$ and non-zero $\vec{d} \in \mathbb{Z}^2$ (Evans, 2003).

These patterns are *emergent* since they are not specified by the rules and only result



(a) Still-life

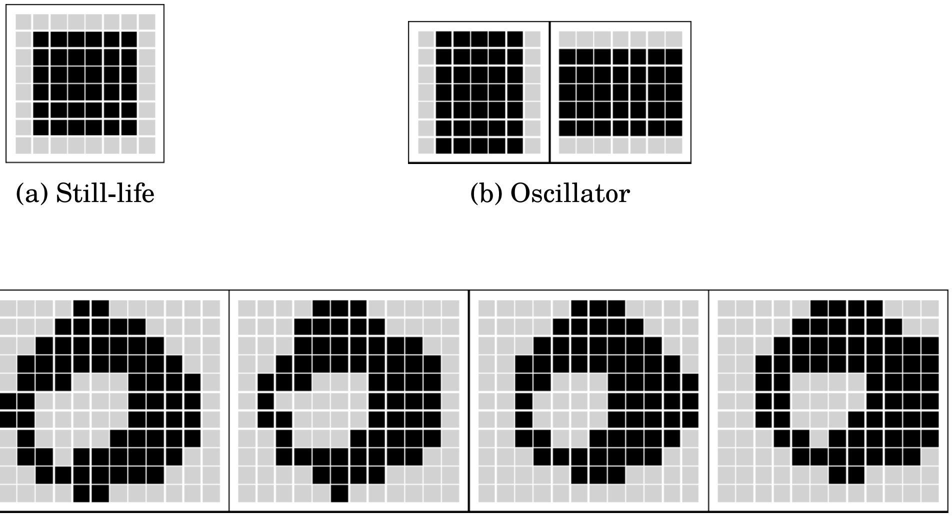
(b) Oscillator

(c) Bug

Figure 2.1: Examples of the three main classes of persistent patterns in GoL. (a) A 2 × 2 *block* with period $\bar{t} = 1$. (b) A 3 × 1 *blinker* with period $\bar{t} = 2$. (c) A *glider* (diagonal bug) with period $\bar{t} = 4$ and translation vector $\vec{d} = \langle 1, 1 \rangle$.

as a consequence of local interactions between cells; they are *precarious* since they can disintegrate when not surrounded by only off-cells; they are *recurrent* by being essentially periodic; and they are *compact* since they consist of a finite set of on-cells all connected via neighbourhood relations. As examples, consider the 2 × 2 *block*, the 3 × 1 *blinker*, and the *glider* of GoL (Figure 2.1). Analogous patterns also exist in other LtL rules with larger ρ. For instance, *Bosco's rule* $(5, \frac{34}{121}, \frac{45}{121}, \frac{34}{121}, \frac{58}{121})$ supports a generalization of the block, blinker, and glider from GoL (Figure 2.2). To emphasize this biological perspective, I will refer to these patterns as *emergent individuals*.

These patterns also appear to scale with larger ρ, approaching a smooth limiting object (Figure 2.3). In fact, Evans (2003) has shown that $(\rho + 1) \times (\rho + 1)$ blocks and $\rho \times (\rho + 2)$ blinkers can be scaled to arbitrary neighbourhood radii. However, no analogous proof yet exists for bugs with stomachs, though empirical findings indicate that they should scale as well. I will reserve a more in depth discussion of scaling for the end of this chapter and the next, but one conceptual point should be made here: pattern scaling should be interpreted as changing not the *size* but the *resolution* of patterns. Concretely, this would amount to mapping $\mathbb{Z}^2$ into $\mathbb{R}^2$ by a factor $(\rho + 4)^{-2}$, fitting more lattice points into the same region of $\mathbb{R}^2$ as ρ increases. This also ensures that the dynamics remain local (the neighbourhood



(a) Still-life

(b) Oscillator

(c) Bug

Figure 2.2: Examples of the three main classes of persistent patterns supported by Bosco's rule. (a) A 6 × 6 block with period $\bar{t} = 1$. (b) A 7 × 5 blinker with period $\bar{t} = 2$. (c) A disoriented bug with period $\bar{t} = 4$ and translation vector $\vec{d} = \langle 2, 1 \rangle$.

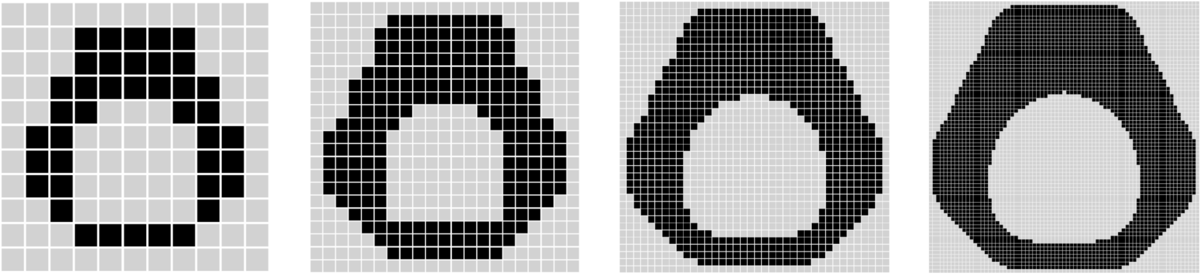


Figure 2.3: Scaling of a translation-invariant bug for $\rho = 5, 10, 20, 40$. Every bug has period $\bar{t} = 1$ and $\vec{d} = \langle 0, n \rangle$ for some n.

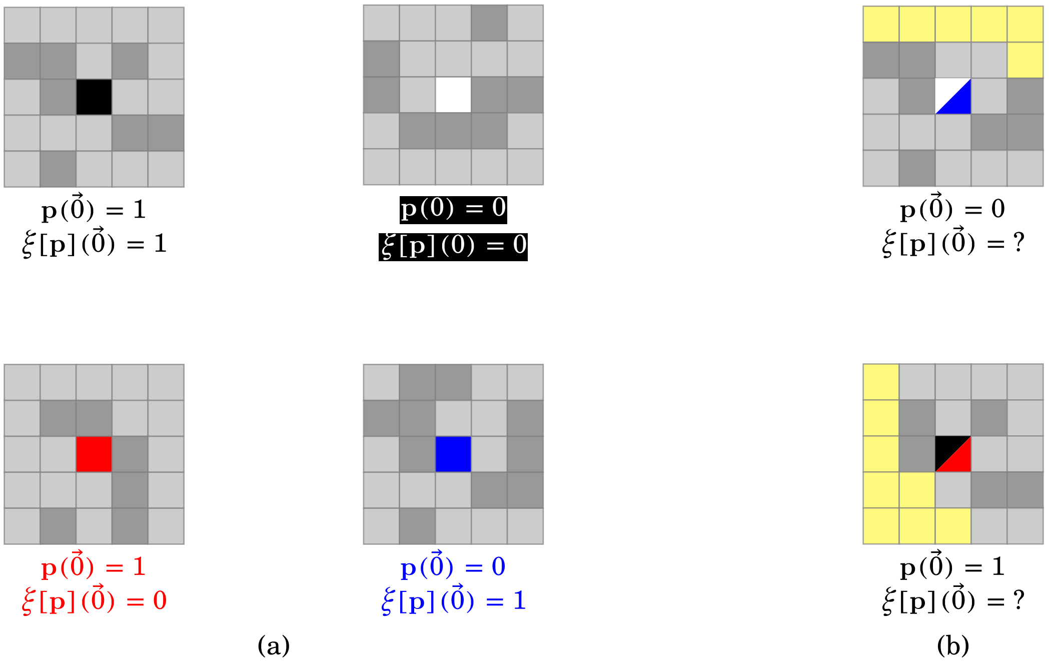
radius remains finite in $\mathbb{R}^2$).

2.2. Interpreting Autopoiesis

To motivate an interpretation of these persistent patterns as autopoietic systems – as something analogous to the living cells of our own universe – it will help to view the unfolding of the LtL universe in *chemical* terms. First, we interpret on- and off-cells as distinct molecular species; let us call them 1- and 0-components, respectively (Beer, 2004, 2015). The update rule, then, describes an interaction between sets of these molecules: the spatial relations between *reactant* components determine a *product* component at the next time step. Of course, this chemistry is very different from that of our universe (for example, a single LtL component participates in multiple reaction simultaneously; components do not persist longer than one time step before transformation, etc.) but the essential relation between reactants and products still holds, and this is sufficient to ground the autopoietic notion of a "process of production." Moreover, the inherent spatiality in this chemistry is very useful from a biological perspective, as we do not have to sacrifice the topological structure of organisms when describing them as networks of processes.

To make this chemical interpretation more concrete, Beer (2015) developed a notation for processes in GoL, which I generalise here to LtL (Figure 2.4a). Every configuration of components in $\mathscr{N}$ determines a process; 0-components are notated in light grey, 1-components in dark grey. These processes fall into four classes: *0-maintenance*, *production*, *1-maintenance*, and *destruction*. 0-maintenance processes, notated in white, occur when the central component in $\mathscr{N}$ is a 0-component and the process produces a 0-component (the cell is off and stays off); production processes are notated in blue (the cell turns on); 1-maintenance processes in black (the cell stays on); and destruction processes in red (the cell turns off).

It will be useful to have a notion of partially determined processes, which are equivalence classes of processes that share the same configuration of components over some subset of $\mathscr{N}$ (Figure 2.4b; Beer, 2020b). More intuitively, a *partial process* is one that does not specify all the components in $\mathscr{N}$, and therefore the type of process (whether 0-maintenance, production, etc.) may be contingent on the state of those unspecified components. I say *may* since LtL allows a partial process to be fully-determined with respect to its type, i.e., when the state of the unspecified components has no bearing on the product of the process. For instance, a GoL process with a central 0-component and five 1-components will always be a 0-maintenance process irregardless of the state of any other components, since the birth interval is $[\frac{3}{9}, \frac{3}{9}]$. I will still notate these processes as if they weren't fully-determined, since this has little bearing on the theory developed later (Section 2.3). All



(a) (b)

Figure 2.4: Processes in LtL for $\rho = 2$. Dark grey cells represent 1-components (on-cells) and light grey cells represent 0-components (off-cells). Every process contains $(2\rho + 1)^2$ cells. Below each process is the state of the center cell and the product of the process. (a) The four main classes of processes. *1-maintenance* processes are notated with a black center-cell; *0-maintenance* processes have a white center-cell; *destruction processes* have a red center-cell; and *production* processes have a blue center-cell. (b) The two main classes of partial processes. *Partial-0* processes are notated with a half-white, half-blue center-cell, representing the possibility of being a 0-maintenance or production processes. *Partial-1* processes are notated with a half-black, half-red center-cell, representing the possibility of being a 1-maintenance or destruction process. Yellow cells have an unspecified state.

partial processes encountered in autopoietic organisations will fall into two classes: partial 0-maintenance/production or partial 1-maintenance/destruction. For shorthand, I will call such processes *partial-0* and *partial-1* processes.

More formally, I define a process as an ordered pair $\psi = (\phi, \mathbf{p})$, where $\phi \subseteq \mathcal{N}$ is the domain of $\mathbf{p} \in \mathcal{U}$. The product of a process is given by $\xi[\mathbf{p}](\vec{0})$, which is undefined when $\phi \neq \mathcal{N}$. In other words, ψ is a partial process when the domain of $\mathbf{p}$ is smaller than $\mathcal{N}$. The density of a process is given by $\Psi(\vec{0}) = (\kappa * \mathbf{p})(\vec{0})$, which is also undefined in partial processes. However, it will be useful to treat this as a lower bound on density by assigning $\mathbf{p}(x) = 0$ for all $x \notin \phi$.

It is important to note that I am making a clear distinction between the formalism of a partial process and its meaning in the context of an organisation. In the formulation presented in Beer (2020b) and Beer, McShaffrey, and Gaul (2024), the unspecified components were always interpreted as specifying the 1-environment. However, as we will see in Section 2.3.1, this need not be the case under my formulation.

To begin moving from a strictly chemical perspective on the LtL universe to a biological perspective, we need to formulate how processes *depend* on or *enable* one another, such that relations of dependency or enablement can form networks. The formulation of processes just described makes this straightforward: a process enables another when its product serves as a reactant. More specifically, a process at site x_1 and time t_1 enables another process at site x_2 and time t_2 if and only if $||x_2 - x_1||_p \leq \rho$ and $t_2 = t_1 + 1$. The converse of this is dependency: a process depends on another when one of its reactants is the product of a preceding process. It will also be important to indicate how a process enables another with respect to how its product participates in the enabled process (Figure 2.5). Thus, the processes underlying an emergent individual form networks of enabling relations that describe the production of the components that constitute that individual. Section 2.3 will define these networks more explicitly.

We now have an intuition for how networks of processes of production can emerge in an LtL universe. But we still need to understand how the notion of a topological boundary dis-

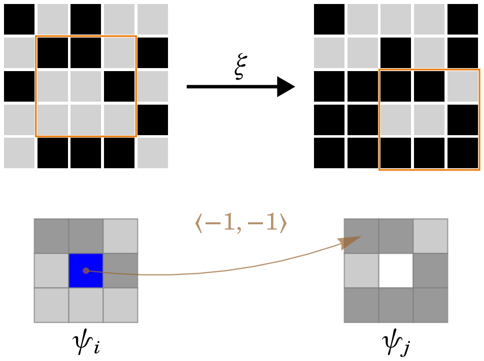


Figure 2.5: Process dependencies in $\rho = 1$ universes. The black and grey grids are toroidal universe states where black cells represent 1-components and grey cells represent 0-components; the one on the right is the state after updating the one on the left once. The orange outlines are the set of cells that realise the process below the universe state. The process on the left, labelled ψ_i, is a production process; the one on the right, labelled ψ_j, is a 0-maintenance process. The brown arrow connects ψ_i to the component it produces in ψ_j, since the location of the center cell of ψ_i aligns with that component after application of ξ. This enabling relation is labelled $\langle -1, -1 \rangle$, the location of the produced component in ψ_j.

tinguishing an emergent individual maps to LtL. In the canonical example of an autopoietic system, the prokaryotic cell, this boundary is a semipermeable lipid bilayer that mediates all interaction with the medium in which the cell exists and encloses its metabolism to prevent diffusion. Traditionally, what counts as a particular pattern in LtL is the specific configuration of on-cells; thus the glider is identified with the on-cells in Figure 2.1c. However, this makes the identification of a boundary impossible since the on-cells do not enclose components. For the same reason, no subset of these components could be said to mediate interaction with the surrounding medium. To remedy this, Beer (2004) interprets the union of the neighbourhoods of the on-cells as components of the individual. This creates a connected ring of off-cells that fully contains the on-cells (Figure 2.6a). Note that this insulates the on-cells from any influence outside the boundary, since all components in the neighbourhood of those cells are either part of the boundary or internal to it – the boundary mediates all interaction with the environment. This union of neighbourhoods interpretation also scales to larger neighbourhood rules (Figure 2.6b).

These components (the on-cells and their neighbourhoods) constitute the *structure* of

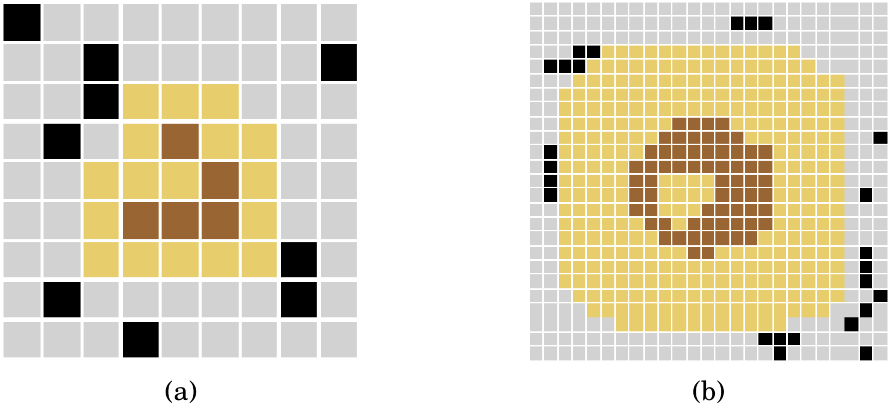


Figure 2.6: Structural realisations of two autopoietic systems. Brown and beige cells are the 1- and 0-components, respectively that realise the individual. Black and grey cells are environmental 1- and 0-components, respectively. (a) A glider (Figure 2.1c) *wedge* configuration in a $\rho = 1$ universe. (b) A bug (Figure 2.2c) configuration in $\rho = 5$ universes.

an individual. Formally, a structure is a pair $(\mathscr{S}, \mathbf{s})$, where $\mathscr{S} \subset \mathbb{Z}^2$ is the set of components that comprise the structure, and $\mathbf{s} \in \mathcal{U}$ is a function describing the state of those components. Picking out a structure also implies an *environment* in which the structure is realised. It will be useful to partition the environment of an individual based what components can influence it. Thus, the *1-environment* is the set of components whose neighbourhood intersects the structure. We will also need to reference the *2-environment* as the set of environmental components outside the 1-environment, but whose neighbourhood intersects it. In general, we can refer to a set $\mathscr{S}$ and its n-environment by a function $\mathcal{E}^n(\mathscr{S})$. Thus, any component in the n-environment of a structure will only be able to influence the processes in $\mathscr{S}$ after n time-steps.

Before proceeding to layout a formal theory in the following sections, some points regarding how I am conceptualising “emergent individual” need to be made. Under the framework employed here, an emergent individual is defined by its autopoietic organisation, such that every realization of that organisation, in a given instance, is a finite configuration of components with a ρ-wide boundary of 0-components. This formulation implies that in any sequence of configurations, every process in one configuration is enabled by a process in the preceding one. It is important to note these points since they reinforce that the identity of an emergent individual is logically determined by its autopoietic organisation. This raises an interesting practical question: how do we get the organisation of an individual? In principle, one could randomly generate networks of processes (defined in Section 2.3) and determine if they are autopoietic (Section 2.4). Such an organisation would uniquely determine a set of possible configurations that realise the network. Of course, this is impractical and does not reflect how we approach biological systems in our universe. A more realistic scenario involves the derivation of an organisation from observed sequences of bounded configurations. However, such a derivation is not, in general, unique: the observed sequences could always be considered a subset of the possible realisations of a larger organisation (Example 7 makes this clearer). This non-uniqueness can be avoided if we assume that the sequences we observe constitute all possible configurations that realise the organisation. One could also stipulate that all configurations in the basin of attraction of a given configuration are realisations of the organisation.

Again, I should emphasise that this is a *practical* problem. It concerns how we map the theory of autopoiesis to empirical observations. It is up to the observer-community to decide how best to apply autopoiesis to describe biological phenomena. All this should make clear that what set of configurations constitute the system is, to a certain extent, arbitrary. For instance, consider a 6×6 block under Bosco's rule (Figure 2.7). It can be perturbed to produce two 1-components from the boundary. If this configuration is left unperturbed – a "null" perturbation, or an environment of all 0-components – it will return to the original block configuration. Should this intermediate configuration be considered an instance of the same system? If so, should all such intermediate configurations be treated the same? What about longer sequences of configurations? One could work out a list of well-motivated answers to questions like these, but the point here is that such stipulations are arbitrary with respect to autopoiesis; it does not matter which network we end up with, only that it be autopoietic. For my purposes, I will only consider the sequence of instances realized in a vacuum, and their symmetries. For example, the glider has two distinct configurations, the *wedge* and the *rocket* (first and second frames of Figure 2.1c, respectively), each with

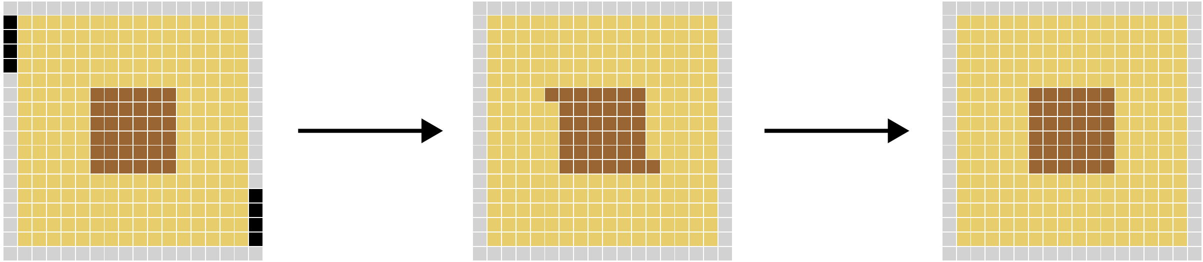


Figure 2.7: A sequence of block configurations in an environment. Arrows represent one application of ξ. Environmental 1- and 0-components are represented by black and grey cells, respectively. Brown and beige cells represent the 1- and 0-components, respectively, that belong to the block. The leftmost configuration is perturbed by the environmental 1-components, resulting in the center configuration with two additional 1-components in the block's boundary. Another update results in the rightmost configuration, an ordinary block in an empty environment.

eight unique symmetries by rotation and reflection. This does not mean that I am only considering vacuum transitions, but just those structures that are observed in a vacuum.

Returning to our interpretation of autopoiesis in LtL more generally, any set of sequences of configurations we stipulate must, taken together, form a closed network of structural transitions: each configuration must, in at least one environment, transition to another configuration in the set, and be transitioned to by some configuration. If such closure is never observed, then the system will not be self-producing, since certain processes that constitute the system depend entirely on processes outside the system (e.g., the first instance of a sequence). I will also require that every component that constitutes an individual in one instance be enabled by a process in the preceding instance. This effectively constrains the possible transitions between structures to those where the structure at $t = 2$ is within the *light cone* of the structure at $t = 1$: every component in the $t = 2$ structure is within ρ of some component in the $t = 1$ structure.

It should be noted that many aspects of this interpretation can be changed besides the criteria for relevant configurations just discussed. For example, one could require that only some processes in a given instance need to be enabled by the preceding instance; consider processes extending beyond the boundary as part of the autopoietic network; allow boundaries to contain 1-components, etc. While some of these changes would fail to be accounted for in the theory presented below, this does not mean that they are necessarily incorrect interpretations of autopoiesis, but that they would require a different mathematical framework. Thus, what I describe below is only one among many possible formulations of autopoiesis in LtL.

2.3. Organisation

What should a mathematical formulation of autopoiesis look like? That is, what mathematical object should be used to describe the organisation of an emergent individual? Most obviously, it needs to be some kind of network whose basic elements are processes. It should also be invariant with respect to changes in how processes are arranged or labelled within the organisation (though this need not be the case for its realisation): only the processes and their dependencies should matter. This means that the object should not explicitly specify how the network is realised; for example, it should not specify which sets of processes co-occur, except insofar as this is implied by dependency relations. Finally, the object should permit the derivation of both the cognitive domain and all possible instantiations of the system.

2.3.1. Process Dependency Graphs

The simplest way to satisfy these desiderata in LtL is with a graph. Following Beer (2015), I define a *process dependency graph* as a digraph with processes as vertices and dependency relations as edges pointing to components in the enabled process. More formally, a process dependency graph is a multidigraph $G = (V, E, \mathfrak{s}, \mathfrak{t})$ (a directed graph permitting parallel edges). Here, V (the vertex set) is an arbitrary finite set representing processes. E (the edge set) is a set of 3-tuples (i, j, k), where $k \in \mathscr{N} \times \mathscr{M}^2$. $\mathfrak{s} : (i, \cdot, \cdot) \mapsto i$ maps edges to their source vertex, and $\mathfrak{t} : (\cdot, j, \cdot) \mapsto j$ maps them to their target vertex.

Processes are not given explicitly in this definition, and are instead implied by the edge labellings in $\mathscr{N} \times \mathscr{M}^2$. To be more specific, edges are triples of a source vertex i (an arbitrary element of V), target vertex j, and an enabling relation $k = (\vec{r}, c_i, c_j) \in \mathscr{N} \times \mathscr{M}^2$. $\vec{r} \in \mathscr{N}$ is the element of ϕ_j (where $\psi_j = (\phi_j, \mathbf{p}_j)$ is the process represented by j) of the component produced by ψ_i, and $c_j \in \mathscr{M}$ is the state of that component; $c_i \in \mathscr{M}$ is the state of the center-component of ψ_i. Each edge, then, specifies how the transformation of a component participates in the organisation.

A process ψ_i can be extracted from the graph by taking the set of edges $\{(\cdot, i, \cdot) \in E\}$, where the state of the $\vec{r}$ component of ψ_i, c, is given by $k = (\vec{r}, \cdot, c)$. Additionally, the $\vec{0}$ component c_0 can be derived from any edge $(i, \cdot, k)$ where $k = (\cdot, c_0, \cdot)$.

I call this formulation of process dependency graphs *edge-centric*, in contrast to the *vertex-centric* formulation present in Beer (2015, 2020b) and Beer, McShaffrey, and Gaul (2024). In the vertex-centric formulation, processes were represented explicitly as vertices – there was no need to recover them from edge-labellings. This has consequences for the interpretation of partial processes. In an edge-centric graph, every reactant of every process must be either produced by another process, or be the center-component of an enabling process. A vertex-centric graph, however, can specify reactants independent of whether they are produced by any other process. Examples 4 and 10 below demonstrate the significance of this difference clearly. I use an edge-centric formulation because it is necessary to connect LtL to RealLife in the next chapter. The theory presented in this chapter, however, will still apply to a vertex-centric formulation and, in some cases, the theory is more powerful there.

Now can we begin exploring the space of autopoietic process dependency graphs. Example 1 is one of the simplest graphs possible in LtL.

> *Example 1: A $\rho = 1$ Block.* Figure 2.8 depicts the organisation of a 2×2 block in universes with $\rho = 1$ (Figure 2.1a). There are 4 1-maintenance processes and 12 partial-0 processes. There are a total of 100 edges.

Note that every specified component of a process in Figure 2.11a (grey and center cells) is produced by another process in the network; that is, processes are only specified with respect to their participation in autopoiesis. This is captured in the above definition by the

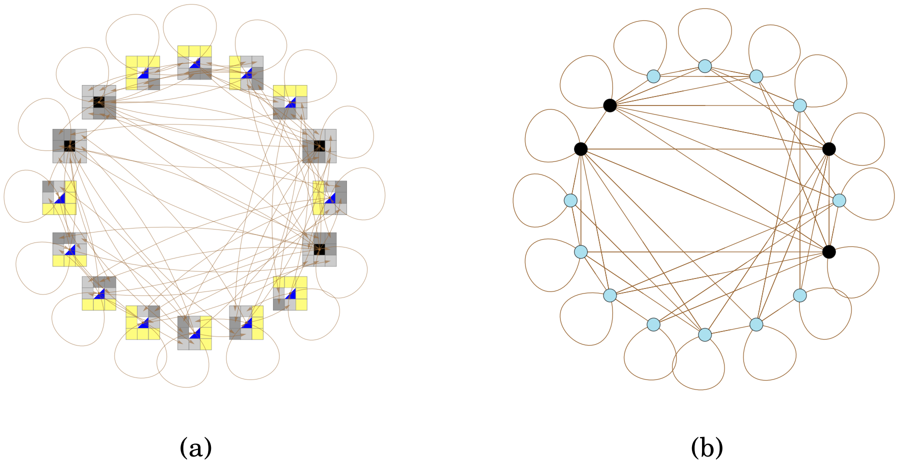


Figure 2.8: The process dependency graphs of a $\rho = 1$ block. (a) Graph with every process represented explicitly, and edges pointing to the enabled component. (b) A reduced representation of the graph where the enabled component is not represented, and antiparallel edges are reduced to undirected edges. Black circles represent 1-maintenance processes, light blue circles partial-0 processes.

edge labels from $\mathscr{N} \times \mathscr{M}^2$, where every edge specifies a component in the target process.

The unspecified components (yellow cells) in Figure 2.8a can only be determined when the organisation is realized in space, since they depend on processes external to the autopoiesis of the block. In the block, partial processes correspond to the boundary, as they mediate interaction with the environment.

Figure 2.8b shows a reduced representation of the process dependency graph in Figure 2.8a. Here, there is no indication of exactly how processes enable each other, only that they do. This graph is also undirected, since all enabling relations are reciprocated in the organisation.

Clearly, this definition of process dependency graphs satisfies the desiderata described above. The processes in Figure 2.8a can be rearranged or labelled in any way, but the graph will remain the same. It also does not specify a particular realization, except insofar as each process specifies its own conditions for realization (the configuration of reactants in the neighbourhood). Lastly, we can derive the cognitive domain and the system's possible instantiations, as will be shown in sections 2.4 and 2.5.

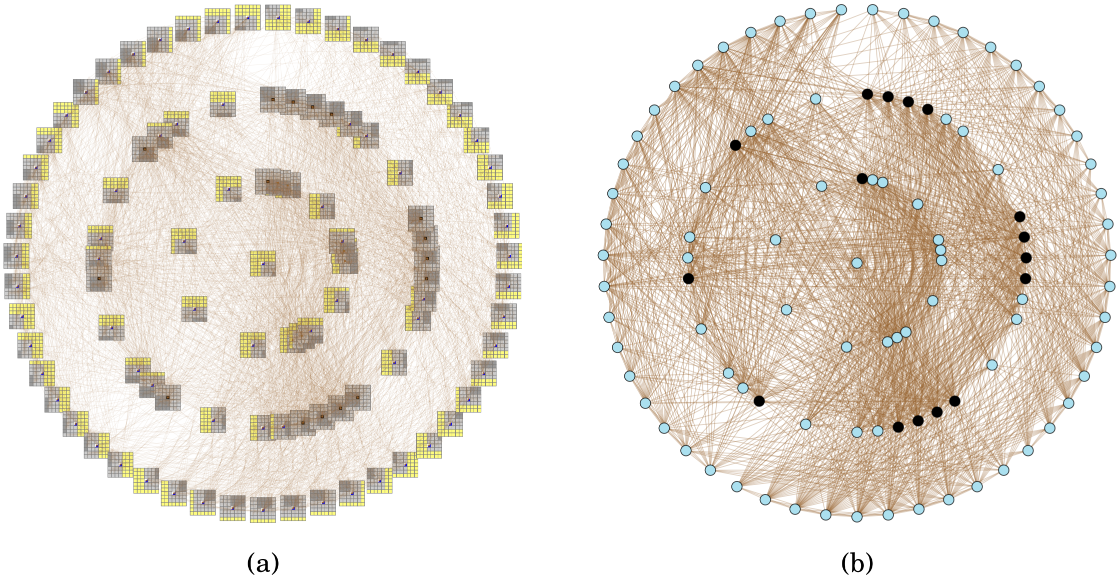
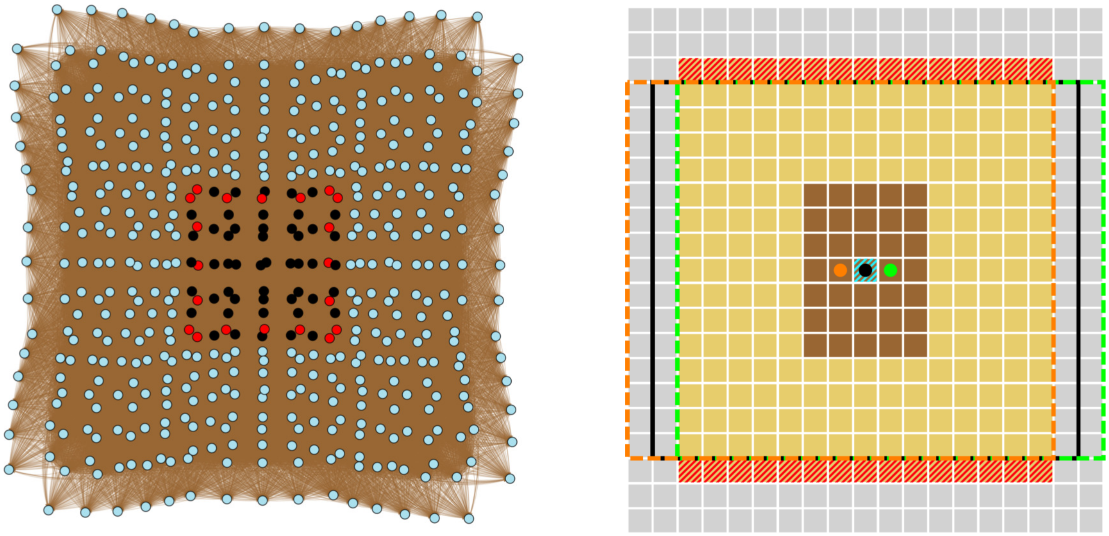


Figure 2.9: The process dependency graphs of a $\rho = 3$ block. (a) Graph with every process represented explicitly, and edges pointing to the enabled component. Every process enables itself, but these edges are not displayed in the shown graph. (b) A reduced representation of the graph where the enabled component is not represented, and antiparallel edges are reduced to undirected edges. Black circles represent 1-maintenance processes, light blue circles partial-0 processes.

> *Example 2: A* $\rho = 3$ *Block.* Figure 2.9 depicts the organisation of a 4×4 block in $\rho = 3$ universes (Figure 2.2a). Out of 100 total vertices, there are 16 1-maintenance processes and 84 partial-0 processes. There are a total of 3,364 edges. Every process in the graph enables itself.

The organization of larger ρ still-lifes is quite similar to the $\rho = 1$ block, but indicates properties common to larger ρ individuals more generally. Consider the $\rho = 3$ block in Example 2. Here, there are partial process whose reactants are mostly unspecified. This implies that larger ρ individuals can generally survive more environmental configurations. It should also be evident that organisations in larger ρ universes are going to have increasing large graphs due to both the number of processes involved and the number of reactants in each one.

> *Example 3: A* $\rho = 5$ *Blinker.* Figure 2.10a depicts the organisation of a $\rho \times (\rho + 2)$



(a) (b)

Figure 2.10: (a) Process dependency graph of a $\rho = 5$ blinker. Black nodes are 1-maintenance processes, red destruction processes, and light blue partial-0 processes. Edges are directed. (b) A set of transitions for a $\rho = 5$ blinker. The black, orange, and green rectangles represent the outlines of the subsequent blinker configurations following different perturbations to the initial configuration (the brown and beige cells). The solid black outline is the transition that occurs when the environment is all 0-components (a vacuum). The cyan-hatched cell corresponds to a 1-maintenance process in the initial configuration. The orange, black, and green circles correspond to the location of the same process under the different perturbations of the corresponding color. Red-hatched cells correspond to processes that don't enable any process in the network.

blinker in $\rho = 5$ universes. Out of 510 total vertices, there are 50 1-maintenance processes, 20 destruction processes, and 440 partial-0 processes. There are a total of 116,910 edges. There are no self-enabling processes.

If we exclude all processes not in the closure and their associated edges, the graph would have 410 vertices with 60 fewer partial-0 processes, and 108,810 edges. I call this subgraph the *closure* of the organisation.

Example 3 describes the organisation of a $\rho = 5$ blinker, an oscillator that demonstrates interesting phenomena not present in the block. The most basic of these differences is that the reduced representation of the organisation is necessarily a directed graph because all transitions of the blinker result in a structural change (a rotation), and thus a new set of processes is realised at every step. More generally, any individual with a period $\bar{t} > 1$ will have this property.

The second interesting property is that there are some processes that do not enable any others: they do not participate in the closure of the organisation. If we look at the transitions of the blinker (Figure 2.10b), there are some components that never overlap with the subsequent structure (the red hatched components). This means that the products of those processes do not constitute the subsequent structure and thus cannot enable any process in it. However, this still satisfies the interpretation of autopoiesis described above: every process in the network is enabled by some other process. The reason I include these processes is that they are still implied by the closure of the organisation. That is, even if they are pruned from the graph, one would recover them in the spatial embedding. Section 2.4 describes the details of this recovery, but the basic idea is that, since these processes are still enabled by the network, all of the reactants that realize them still are produced. Thus, one could just as well treat the closed network as the proper representation of the organisation.

The organisation is also a *multidigraph*: there is more than one edge in the same direction between some pairs of vertices. This arises because there are structural transitions that differ only by translation. Figure 2.10b shows the three transitions a vertical blinker structure can undergo. The transition that occurs in a vacuum (black outline) results in the subsequent structure centred on the initial structure. The other two non-vacuum transitions result in displacements of one cell in either direction horizontally (orange and green outlines). By symmetry, the same set of transitions apply in the orthogonal direction. That this results in a multigraph is clear if one considers 1-maintenance processes. Say a process ψ_i in the vertical blinker (cyan-hatched cell in Figure 2.10b) enables the $\langle 0, 0 \rangle$ component of a process ψ_j in the horizontal blinker under a null perturbation (black circle). If a different perturbation causes the horizontal blinker to instead move right, then the relative position of ψ_i and ψ_j changes (ψ_j is now the green circle), which means ψ_i would enable the $\langle -1, 0 \rangle$ component of ψ_j. Thus, there is more than one edge from ψ_i to ψ_j in the organisation.

Blinkers are also the simplest class of individuals that *shed* and *consume* components. That is, in any given structural transition, the blinker will produce components that do not constitute the individual at the next step (e.g., the products of the red-hatched processes in Figure 2.10b); conversely, it will be constituted by components produced by environmental processes (where the subsequent structure overlaps with the environment in Figure 2.10b). In this way, autopoietic systems in LtL express a dependence on the environment to constantly renew themselves. More concretely, this means that not every process will participate in any given transition.

> *Example 4: A $\rho = 1$ Blinker.* Figure 2.11a depicts the organisation of a blinker in $\rho = 1$ universes. Out of 30 total vertices, there are 2 1-maintenance processes, 4 partial-1 processes, and 24 partial-0 processes. There are a total of 306 edges. All processes are in the closure of the graph. The organisation is represented as a multigraph, but results from a different set of transitions (Figure 2.11b) than the $\rho = 5$ blinker. The organisation is equivalent to its closure. All circled processes in Figure 2.11a have no edges pointing to their center-component.

Example 4 introduces processes whose center-components are *always* dependent on environmental processes. This is in contrast to Example 3 where every center-component is enabled by another process in the organisation in at least one transition. One consequence of this is that there will be partial processes that do *not* correspond to the boundary of the blinker, i.e, partial-1 processes. Thus, the meaning of an unspecified reactant becomes more nuanced: *where a reactant is unspecified, the enabling relation corresponding to its production is external to the organisation*. Put another way, an unspecified reactant means either (i) the component realising that reactant is an environmental component, or (ii) the process producing that reactant is not part of the organisation, or (iii) both. However, these reactants can be fully-determined in the spatial embedding of the organisation when the center-component of an adjacent process is known, which is true for all processes in the organisation – in other words, the reactant can be determined when (ii), but not (i), is true. (Section 2.4 will explain in more detail why this is the case.) Note again that this differs

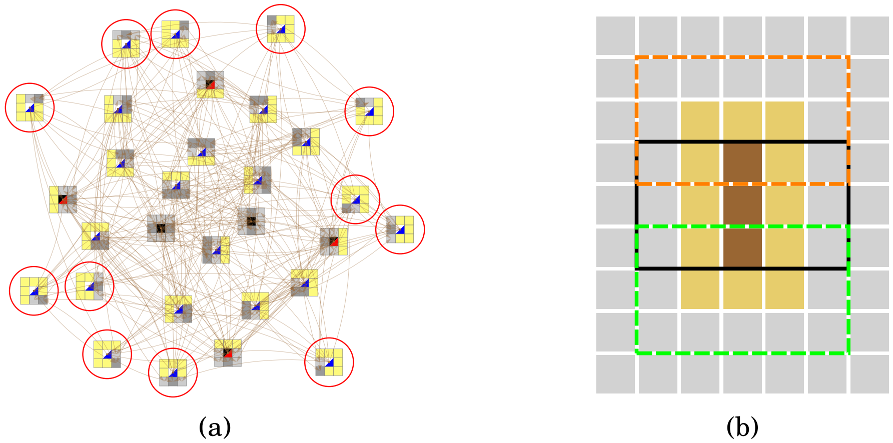


Figure 2.11: (a) Process dependency graph of a $\rho = 1$ blinker. The circled processes have no edges pointing to their center-component. (b) A set of transitions for a $\rho = 1$ blinker. The black, orange, and green outlines represent the bounds of the subsequent blinker configuration following different perturbations to the initial configuration (the brown and beige cells). The black outline corresponds to the transition that occurs when the initial configuration undergoes a null-perturbation.

from the interpretation of partial processes used in Beer's work. There, an unspecified reactant meant only that the reactant, when realised, constitutes the 1-environment of the system. In contrast, an unspecified reactant here implies that there is no edge in the graph pointing to it.

Example 4 brings attention to another important aspect of the formalisation of process dependency graphs described above. Specifically, the edge labels need to specify the center-component of the source process precisely because there are processes that are never enabled in that way. This blinker also demonstrates that what components are shed in a given transition need not be restricted to the products of boundary processes. For example, the center-most process of the blinker in Figure 2.11b (a 1-maintenance process), only produces a component that constitutes the blinker in one transition (the black outline). In the other two transitions (orange and green) the product of that process becomes part of the blinker's environment.

Example 5: A $\bar{t} = 4$ $\rho = 5$ Bug. Figure 2.12 depicts the organisation of a $\bar{t}$ disori-

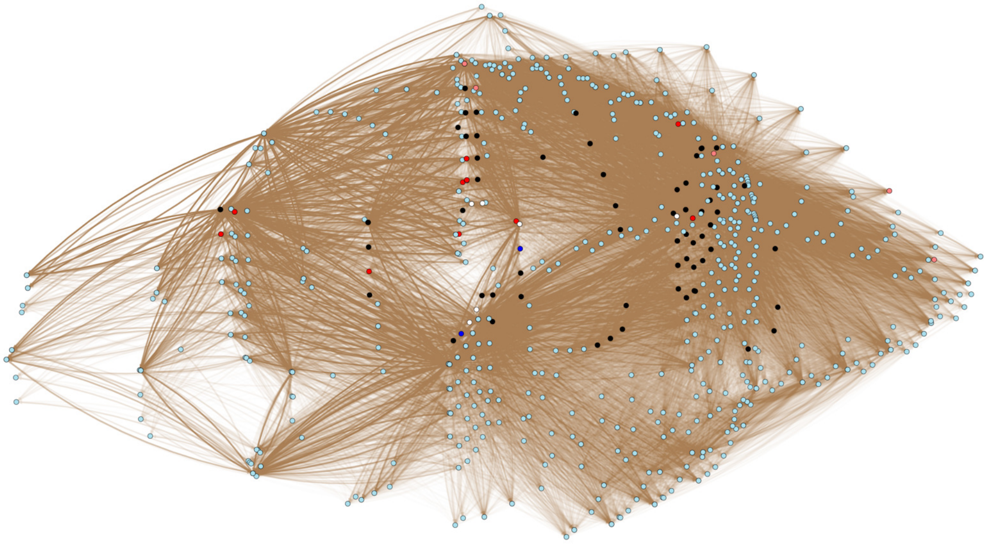


Figure 2.12: Process dependency graph of a $\rho = 5$ 4-cycle disoriented bug. Black nodes are 1-maintenance processes, white nodes are 0-maintenance processes, red nodes are destruction processes, blue nodes are production processes, pink nodes are partial-1 processes, and light blue nodes are partial-0 processes. Edges are directed.

ented bug in $\rho = 5$ universes (Figure 2.2c). Out of 1,539 total vertices, there are 50 0-maintenance processes, 197 1-maintenance processes, 8 production processes, 31 destruction processes, 23 partial-1 processes, and 1,230 partial-0 processes. There are a total of 137,323 edges. The closure of the graph has 1486 vertices and 134,292 edges, with 53 fewer partial-0 processes.

For any given process, there is at most one edge labelled with any given $\vec{r} \in \mathscr{N}$. Such an organisation is called a *reversible organisation*.

Example 5 is the first non-still life *reversible* organisation described so far – an organisation in which every component of every process has only one edge pointing to it. This arises because the realisations of the organisation are restricted to the sequence of configurations in Figure 2.2c. This implies the existence of 7 other symmetric organisations by rotation and reflection of the structures. This points to a limitation of assuming only vacuum structures and their symmetries: as ρ increases, there are more processes isolated from the environment and thus less flexibility in how the morphology of a bug can change.

Looking to the "stomach" of the bug in Figure 2.2c, all of those processes and the ones immediately surrounding are fully-determined. It is therefore impossible for the morphology of the stomach to change to something other than the next one in the shown sequence. Any non-reversible generalisation of Example 5, then, must include sequences in which the stomach of a configuration can only change after at least two time-steps by first changing the state of the surrounding partial processes.

> *Example 6: A Translation-Invariant $\rho = 5$ Bug.* Figure 2.13 depicts the reversible organisation of a translation invariant bug in $\rho = 5$ universes with a translation vector $\vec{d} = \langle 0, 1 \rangle$ (Figure 2.3, first frame). Out of 343 vertices, 17 are 1-maintenance processes, 18 are 0-maintenance processes, 5 are production processes, 14 are destruction processes, 9 are partial-1 processes, and 280 are partial-0 processes. All processes are self-enabling, except for 19 partial-0 processes. There are a total of 29,493 edges. The closure of the graph has 324 vertices and 28,419 edges, with 19 fewer partial-0 processes.

Translation-invariant bugs, such as Example 6, are bugs that move without any modification to their structure. In Evans'(2003) formalism, they are functions $\mathbf{u} \in \mathcal{U}$ such that $\xi[\mathbf{u}] = \sigma^{\vec{d}} \circ \mathbf{u}$ for some $\vec{d} \in \mathbb{Z}^2$. These bugs are interesting for us primarily because they are the subclass of bugs whose scaling has been investigated most thoroughly (Evans, 2003), while still exhibiting non-trivial morphologies (i.e., non-rectangular). Thus, they will serve as the most complex example in scaling organisations to RealLife in Chapter 3.

> *Example 7: The Glider.* Figure 2.14 depicts the organisation of the glider (a $\rho = 1$ diagonal bug; previously computed in Beer, McShaffrey, and Gaul, 2024). Out of 352 total vertices, there are 16 0-maintenance processes, 48 1-maintenance processes, 32 destruction processes, and 256 partial-0 processes. There are a total of 5,912 edges. There are no self-enabling processes. The closure of the graph has 320 vertices and 5,608 edges, with 32 fewer partial-0 processes.

The glider is the most well-studied individual in LtL (Beer, 2004, 2014, 2015, 2020a,

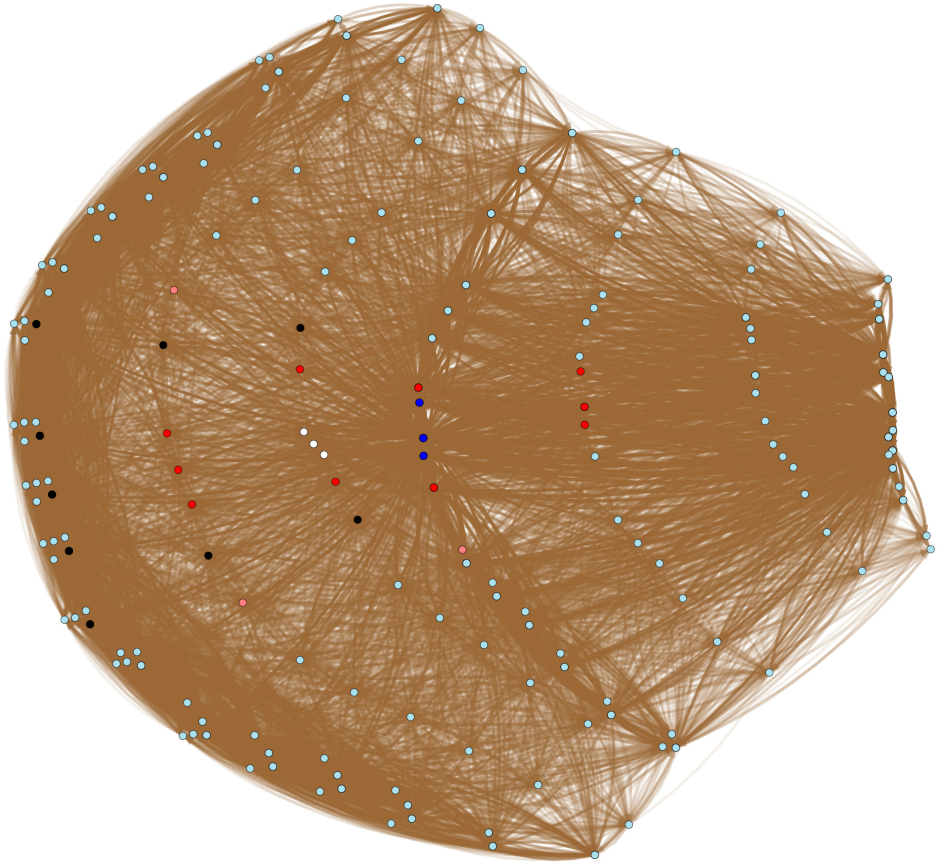


Figure 2.13: Process dependency graph of a $\rho = 5$ translation invariant bug. Black nodes are 1-maintenance processes, white nodes are 0-maintenance processes, red nodes are destruction processes, blue nodes are production processes, pink nodes are partial-1 processes, and light blue nodes are partial-0 processes. Edges are directed.

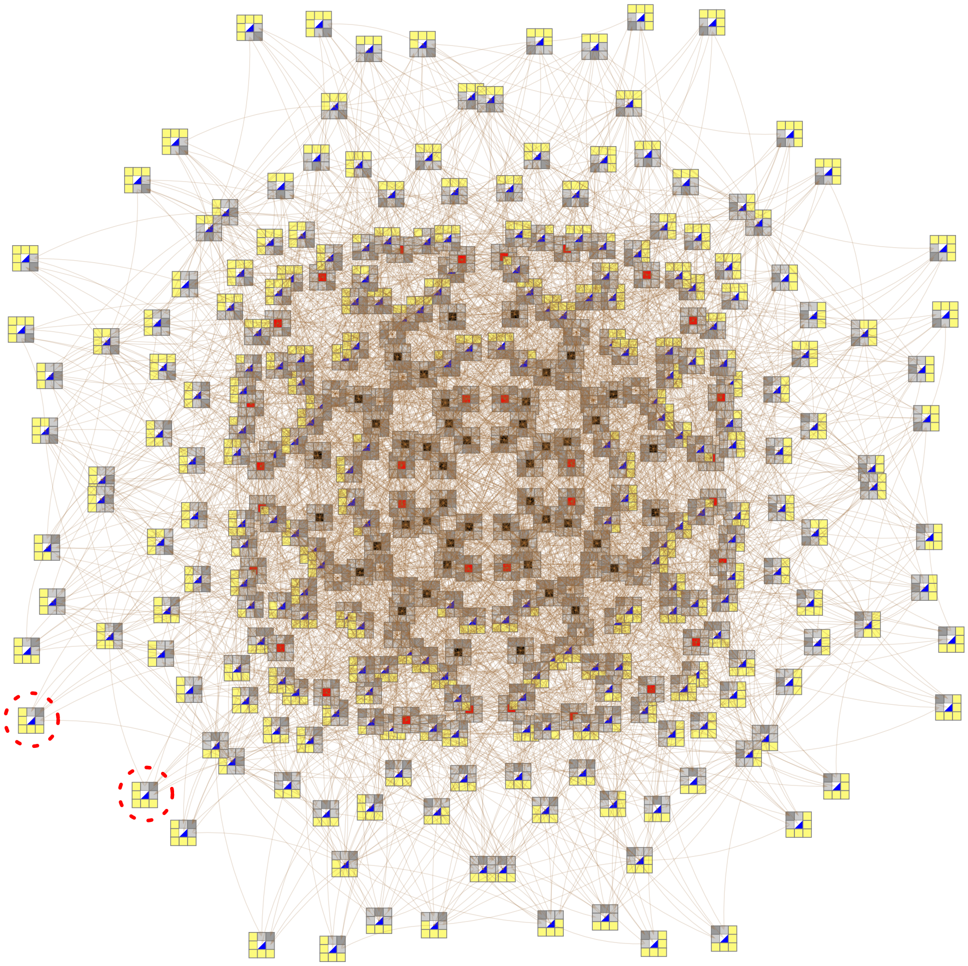


Figure 2.14: Process dependency graph of a the glider (Example 7). The two circled processes are *chemically-equivalent* (are described by the same $(\phi, \mathbf{p})$). (This figure is a reproduction of Figure 2 from Beer, McShaffrey, and Gaul (2024, p. 25))

2020b, 2020c; Beer, McShaffrey, & Gaul, 2024). It is the simplest autopoietic system that exhibits movement, but also possesses the most complex cognitive domain of any individual discussed so far (Figure 2.19c). Thus, I will use it as the main example in Section 2.4 when deriving the cognitive domain, especially as it is easier to visualise than the other bugs.

The glider also makes clear two interesting properties. First, its organisation contains four symmetric reversible organisations. Looking at Figure 2.19c, these subgraphs describe the four 4-cycles with black and gray edges. This contrasts with Example 5, where the symmetries do not form a larger organisation.

It also brings out most clearly a point discussed in Section 2.2: the non-uniqueness of observation in determining an organisation. For example, if we only observe one of the glider's 4-cycles, we could adequately describe its organisation by the corresponding subgraph of Example 7. However, those observations could just as well fit with the full organisation, where only a subset of the its possible realisations are ever observed. Thus, how one decides to relate observations to organisation will depend on how exhaustive one takes their observations to be, and precisely how limited sets of observations are related to the (potentially) broader space of an individuals interactions and structural realisations. Here, I take observations to be exhaustive of structural realisations up to rotational and reflectional symmetries, and then derive all possible transitions among those structures (under a given LtL rule) to determine the full space of an individual's interactions.

The glider also makes clear that apparently identical processes can be distinct with respect to organisation. For example, consider the two circled processes in Figure 2.14. Both are described by the same $(\phi, \mathbf{p})$, but are distinguished by how they participate in the organisation – what processes they enable and depend on – and thus they correspond to different labels in V. I call these processes *chemically equivalent*, but organisationally distinct.

2.4. The Cognitive Domain

In the previous section, I made frequent reference to the structural transitions and spatial embeddings of an organisation, without any explanation regarding precisely how these are all related. This section fills that gap by showing how an *interaction graph* – a network describing the possible transitions between an organisation's spatial realisations – can be derived from an organisation without any additional assumptions.

Before proceeding to the derivation procedure, we will need to be more precise about what an interaction graph should look like. As Beer (2020a) shows, there are many possible formulations of the interaction graph, depending on how structural symmetries (translation, rotation, and reflection) are treated. However, we can begin narrowing the space of formulations by considering the desiderata we would want such a formulation to satisfy. First, it would be useful if, given an initial spatial embedding of the organisation and a perturbation, the interaction graph uniquely determines the subsequent structure *and* its spatial embedding. An interaction graph that, for example, only specified transitions between a glider's *rocket* and *wedge* structures would therefore not satisfy this criterion, as it could not distinguish between perturbations that resulted in different symmetries and spatial embeddings of those structures.

We will also want interaction graphs to abstract over translational symmetries – otherwise identical *rocket* configurations should not be treated as distinct realisations of the glider based on their position in space. Thus, an interaction graph should only specify *relative* translation between structures. This also means that the spatial embeddings derived from an organisation should not specify the absolute position of any process in space, but rather how processes are positioned relative to each other.

Finally, an interaction graph should group perturbations that result in an identical transition between pairs of structures. This is to emphasise the perspective of the individual over that of the observer, since the differences between such perturbations are not relevant to the persistence of the organisation. Such groupings are called *perturbation* *classes*.

Other desiderata could be specified, but any formalisation that satisfies the above should only differ in how useful they are practically. For instance, one could separate different symmetries of a structure into distinct vertices of the interaction graph, or encode different transformations (reflection, rotation, etc.) into edges between a smaller set of structures. I will take the former approach, whereby every edge in the interaction graph is a unique structure up to translation, and edges encode relative translation between pairs of structures.

More formally, I define an *interaction graph* $\mathcal{I}$ as an edge-decorated multidigraph $(\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$, where $\mathcal{S}$ is a set of *structures* $S = (\mathscr{S}, \mathbf{s})$, $\mathcal{T}$ is a set of transitions $(S_i, S_j, \vec{v})$, $\mathfrak{s} : (S_i, \cdot, \cdot) \mapsto S_i$ is a map from edges to the source vertex, $\mathfrak{t} : (\cdot, S_j, \cdot) \mapsto S_j$ is a map from edges to the target vertex, and $\tau : \mathcal{T} \to \mathcal{C}$ is a map that labels edges with elements from an arbitrary finite set $\mathcal{C}$ representing the perturbation classes.

In order to derive an interaction graph from a process dependency graph, we need to get a set of *spatial embeddings* that map processes to $\mathbb{Z}^2$. From these embeddings, we can attain both the structures and the edges of the interaction graph.

The basic logic of this derivation is grounded in the spatiality of enabling relations and the LtL chemistry. For example, consider again Figure 2.5. Here, the spatial displacement of sequential processes implies a specific enabling relation between those processes. But this logic also works in reverse: a given enabling relation implies a relative spatial displacement. We can therefore take the set of edges emanating from a given process and derive a set of embedding fragments (Figure 2.15).

The procedure for deriving these partial embeddings is as follows (outlined in Figure 2.15). First, find the most frequently enabled reactant (the corresponding edges highlighted in magenta). Then, place the enabled processes in a $(2\rho + 1) \times (2\rho + 1)$ grid in the spot corresponding to its spatial displacement from the enabling process (assumed to be centered on the grid). Further processes can be added to the grid if two conditions are satisfied: (i) the union of the process functions is well-defined (Figure 2.16), and (ii) the

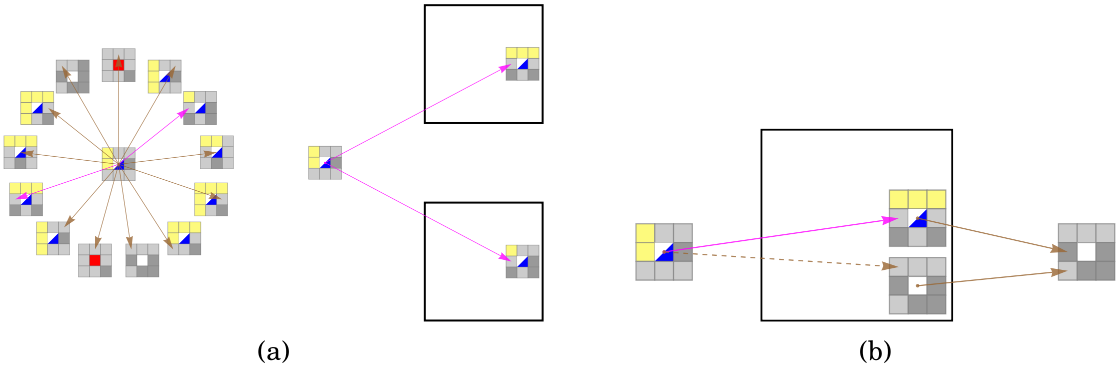


Figure 2.15: Steps in the extraction of embedding fragments from a single process. (a) Separation of fragments. Brown edges are enabling relations as before, and magenta edges are enabling relations that point to the most frequently pointed to component: edges $(i, j, (\vec{r}'_0, c_i, c_j))$ where $\vec{r}'_0$ is the most frequent element of all $\vec{r}$ emanating from the central process. The processes enabled by the magenta edge are separated into two fragments (black boxes), and placed at $-\vec{r}'_0$. (b) Adding processes to a fragment. The process enabled by the dotted edge can be added to the fragment at $-\vec{r}_0$ because it enables another process by $\vec{r}_1$, and the process at $\vec{r}'_0$ enables the same by $\vec{r}'_1$, and $\vec{r}_0 - \vec{r}'_0 = \vec{r}_1 - \vec{r}'_1$.

candidate process being added enables the same process as another in the grid, and the relative position of the processes in the grid is the same as the relative position of the reactants they enable. Finally, all remaining processes that have not been integrated into a partial embedding can be added through the same procedure, except condition (ii) is ignored.

The union of process functions is defined by a set of processes $\{\psi_i\}$ and a spatial embedding $f : \{\psi_i\} \to \mathbb{Z}^2$:

$$\mathscr{S} = \bigcup_i \phi_i + f(\psi_i) \tag{2.5}$$

$$\mathbf{s} = \bigcup_i \sigma^{f(\psi_i)} \circ \mathbf{p}_i, \tag{2.6}$$

where $\mathbf{s}$ is well-defined when no processes conflict over the intersection of their (translated) domains. Figure 2.16 shows an example of how process embeddings specify a universe state function.

Once partial embeddings have been extracted from every process in the organisation, they can be merged by aligning identical processes (Figure 2.17). More specifically, given

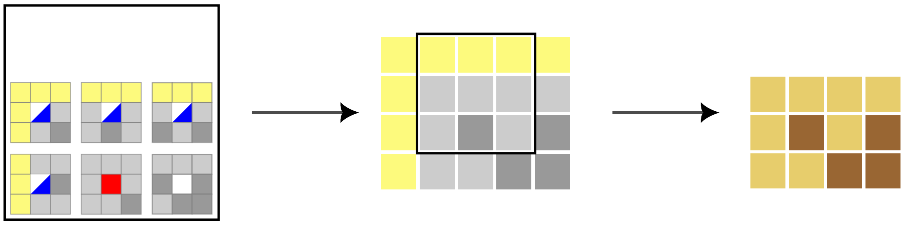


Figure 2.16: Structure from process function agreement. On the left is a spatial embedding of processes. In the middle is the union of the process functions from that embedding. The black box corresponds to the box in the embedding, except each process is replaced by its center-component. On the right is the union of functions expressed in the structure notation, where unspecified components are removed.

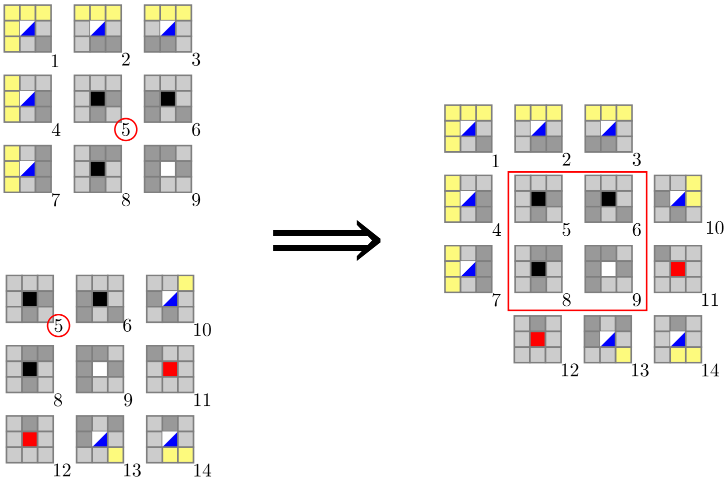


Figure 2.17: Merging of embedding fragments. On the left are 2 process embeddings, where each process is labelled by an index. The fragments contain the same process (labelled 5), and are therefore merged. On the right is the merged embedding. Boxed in red are all the processes that are shared by the fragments on the left.

the set of process labels V from the process dependency graph, if two partial embeddings contain a process with the same label, all processes in one of the embeddings are shifted so that the identically labelled processes have the same embeddings ($f(\psi_i) = f'(\psi_i)$). Then the two embeddings are combined, dropping any redundant processes. When no two embeddings can be merged, the full spatial embeddings of the organisation have been derived.

It is then a straightforward procedure to get the interaction graph $\mathcal{I}_G = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ from the spatial embeddings $\{f_i\}$ and the process dependency graph $G = (V, E, \mathfrak{s}, \mathfrak{t})$. The

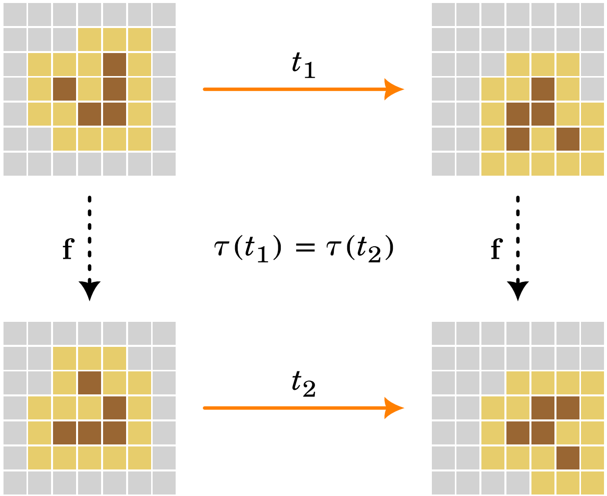


Figure 2.18: Matching edge labels from symmetry. On top is a transition described by the edge $t_1 \in \mathcal{T}$, labelled ORANGE. On the bottom is another transition described by the edge $t_2 \in \mathcal{T}$. **f** is a transformation that rotates universe state functions by $-\frac{\pi}{2}$ and then reflects them across the vertical axis. Since **f** can be applied to both structures in t_1 to get the structures in t_2, t_2 is also labelled ORANGE ($\tau(t_1) = \tau(t_2) =$ ORANGE).

structure set $\mathcal{S}$ can be attained by taking the union of the process functions for each embedding – resulting in a set of structures $S = (\mathscr{S}, \mathbf{s})$ – and normalising them (centering about $\vec{0}$). The edge set $\mathcal{T}$ can be attained by finding all unique pairs $(S_1, S_2, \vec{v})$ where ψ_i and ψ_j are processes in S_1 and S_2, respectively, and $(i,j,k) \in E$, with $k = (\vec{0}, \cdot, \cdot)$; each such edge is then given by $(S_1, S_2, f_2(\psi_j) - f_1(\psi_i))$, which specifies a translation vector between S_1 and S_2.

Deriving the perturbation class function $\tau : \mathcal{T} \to \mathcal{C}$ is more involved. The basic idea is to assign pairs of edges the same label in $\mathcal{C}$ if the same transformation (rotation and/or reflection) can be applied to the source and target structure of one edge to get the other (Figure 2.18). More specifically, if $t_1 = (S_{1i}, S_{1j}, \vec{v}_1)$ and $t_2 = (S_{2i}, S_{2j}, \vec{v}_2)$ are edges in $\mathcal{T}$, then $\tau(t_1) = \tau(t_2)$ when there exists some transformation **f** such that $\mathbf{f} \circ \mathbf{s}_{1i} = \mathbf{s}_{2i}$ and $\mathbf{f} \circ (\sigma^{\vec{v}_1} \circ \mathbf{s}_{1j}) = \sigma^{\vec{v}_2} \circ \mathbf{s}_{2j}$.

The following examples will clarify aspects of the above definitions and demonstrate some of their consequences.

Example 8: The Glider. Figure 2.19a shows the spatial embeddings of the glider organisation (Example 7). Note that there are 16 distinct embeddings here, and that their placement in the figure is arbitrary: only the positions of processes *within* an embedding is meaningful. What is meaningful between different embeddings, however, are the bundles of edges connecting them. Each of these bundles describes the dependency relations involved in transitions between glider structures. We can see this more clearly by labelling the edges according to the symmetry consideration discussed above. Finally, we can replace each embeddings with the structure it implies (Figure 2.16) and collapse the bundles into labelled edges between structures. This gives the *interaction graph* of the glider (Figure 2.19c).

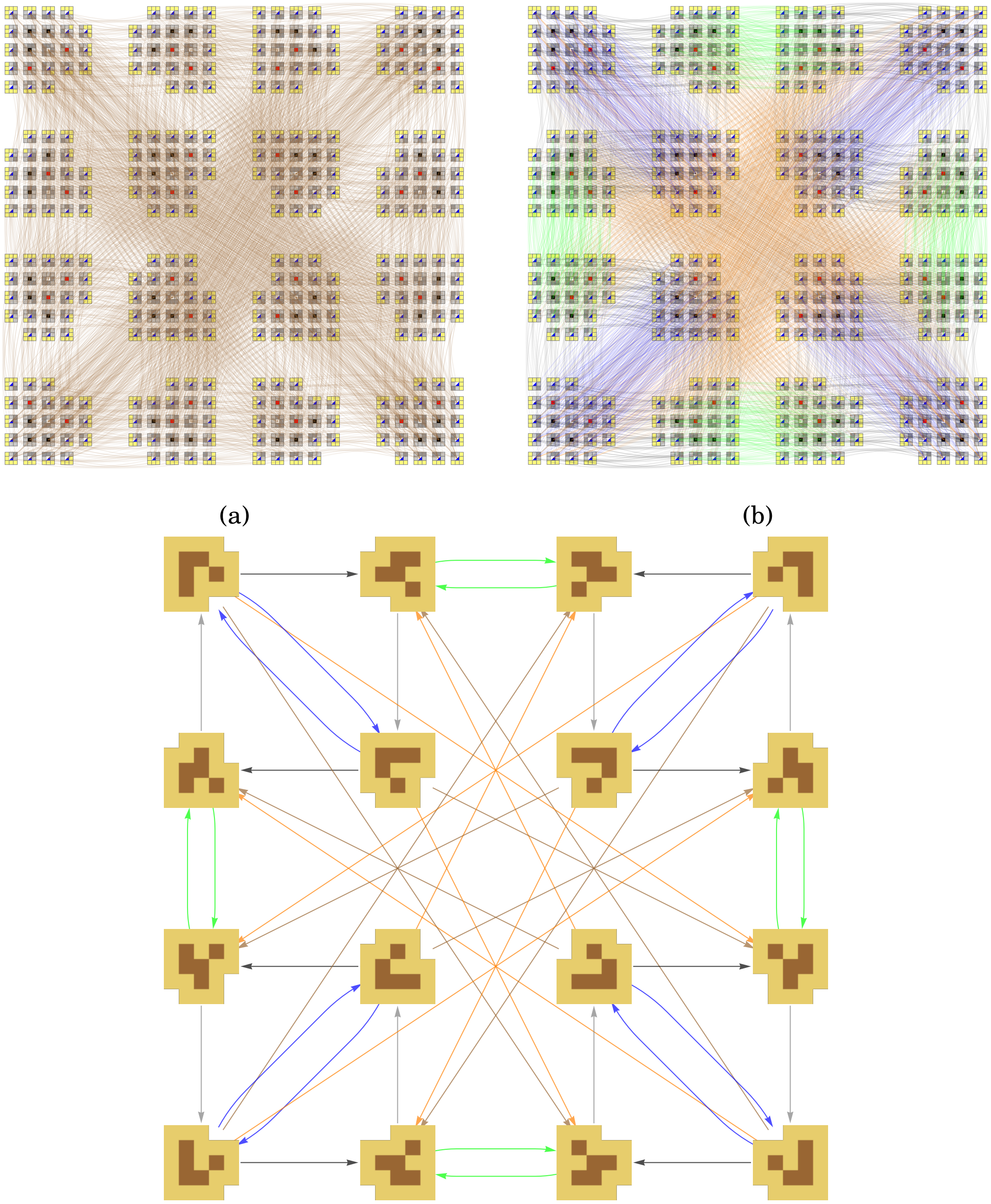
The interaction graph has 16 distinct spatial embeddings of the organisation, realising two basic structures (*wedge* and *rocket*) with 8 symmetries each. There are 48 edges divided into 6 non-destructive perturbation classes:

$$\mathcal{C} = \{\texttt{BLUE}, \texttt{ORANGE}, \texttt{BROWN}, \texttt{BLACK}, \texttt{GRAY}, \texttt{GREEN}\}\,.$$

The BLACK and GRAY classes are equivalent to null-perturbations.

This interaction graph can be used to predict the glider's response to any perturbation. For example, all possible perturbations ((1,2)-environments) on a wedge structure can be partitioned into 5 sets: BLUE, ORANGE, BROWN, BLACK, or destructive. Destructive perturbations destroy the glider organisation, while the other four result in some transition to another structure. The edges of the graph will tell us how that subsequent structure is positioned relative to the initial wedge. — More generally, any observed transitions between glider structures corresponds to a unique edge in the interaction graph.

Another useful property of the graph is that each basic structure (wedge and rocket) has a unique set of perturbation classes that is preserved across all symmetries. This means that, no matter how the glider is embedded in space, its internal state determines the space of its possible behaviours. An important conceptual point here is that internal state is not



(a)

(b)

(c)

Figure 2.19: Interaction graph from the glider organisation. (a) The 16 spatial embeddings of the glider organisation with edges retained. (b) The spatial embeddings with edges labelled according to the interaction graph. (c) The interaction graph of the glider, with 6 perturbation classes in black, gray, blue, green, brown, and orange. (These figures are reproductions of Figures 4 5A and 5C from Beer, McShaffrey, and Gaul (2024, pp. 27–28).)

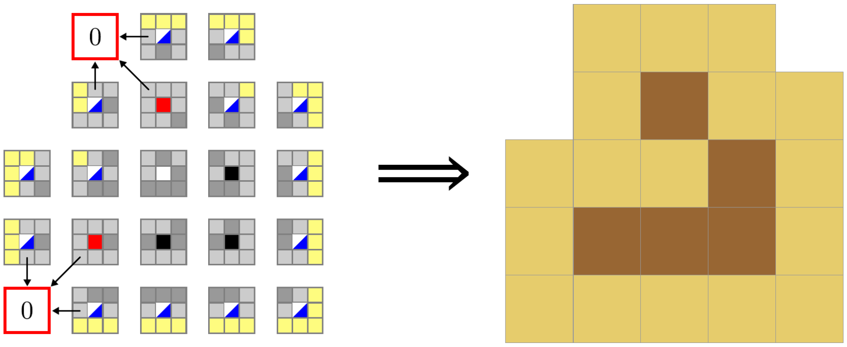
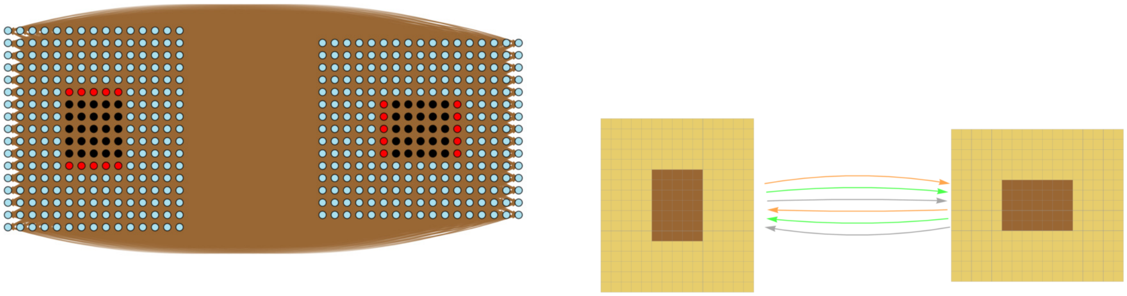


Figure 2.20: Glider structure from the closure of the glider organisation. On the left is one of the 16 spatial embeddings of the closed glider organisation. The 0s boxed in red represent processes with a 0 central-component. They are implied by the reactants pointing to them. On the right is the structure attained by replacing each process with its central-component, and replacing the boxed 0s with 0-components.

sufficient to specify how a system is structurally coupled to its environment. However, it is sufficient to fully specify the effects of all possible perturbations *from the perspective of the system*. This is why the interaction graph contains both labelled perturbation classes and distinguishes symmetric structures: to capture both the perspective of the system and its possibilities for structural coupling. Relatedly, the interaction graph can be interpreted as a nonautonomous dynamical system, where the state space is {glider, wedge}, and the nonautonomous inputs are the perturbation classes.

The derivation process also works for the closure of the organisation. Figure 2.20 shows how an embedding of the closed glider organisation still implies a proper structure (i.e., with a full boundary) despite missing two processes from the non-closed organisation. It should be clear that, despite not representing the missing processes explicitly, they are still implied in any realisation of the glider by the neighbouring processes – all processes that satisfy the conditions implied by those neighbours fall into the equivalence class represented by the process in the non-closed organisation. This is what grounds the claim made in Section 2.3.1 that a closed organisation is just as valid of a representation as the non-closed form.

Example 9: A $\rho = 5$ Blinker. The interaction graph of the $\rho = 5$ blinker in Example 3 (Figure 2.21b) has 2 distinct spatial embeddings of the organisation



(a) (b)

Figure 2.21: (a) The 2 spatial embeddings derived from the $\rho = 5$ blinker organisation, with edges retained. (b) The interaction graph derived from the $\rho = 5$ blinker organisation, with 3 perturbation classes in orange, green, and gray.

– *horizontal* and *vertical* – realising 1 basic structure with 2 unique symmetries. There are 6 edges divided into 3 non-destructive perturbation classes:

$$\mathcal{C} = \{\texttt{GRAY}, \texttt{ORANGE}, \texttt{GREEN}\}\,.$$

The `GRAY` class is equivalent to a null-perturbation. The interaction graph is a multidigraph.

The reader should recall that the organisation of the $\rho = 5$ blinker was also a multigraph. This hints at a more general pattern: a dependency multigraph implies an interaction multigraph. In the organisation, parallel edges represented different enabling relations, but in the interaction graph, parallel edges represent different translations of the target structure relative to the source structure. The logic behind this is rather straightforward. First, every edge must correspond to one transition and every process to only one structure. And since parallel edges cannot be realised simultaneously, there must be more than one transition between the structures that the involved processes belong to – the interaction graph must be a multigraph.

The blinker also makes even more explicit the distinction between internal state and spatial embedding: the former is constant, whereas the latter varies by both the symmetry

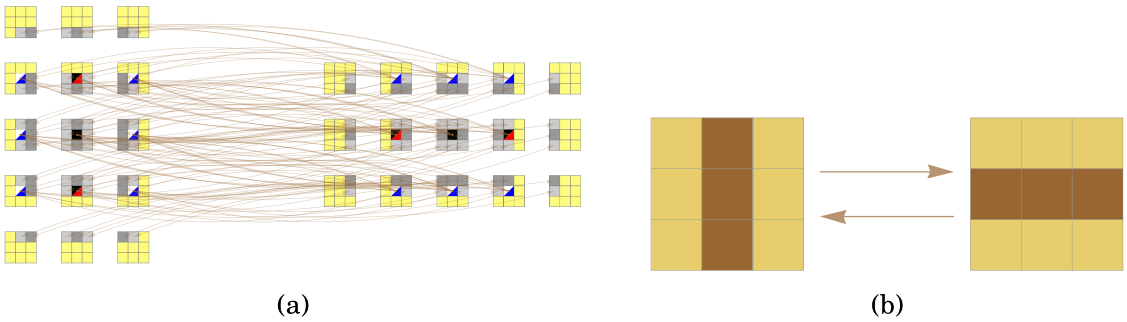


Figure 2.22: (a) The 2 spatial embeddings of the reversible blinker organisation, with edges retained. (b) The interaction graph derived from the reversible blinker organisation.

of the basic structure and the different relative translations in any given transition.

> *Example 10: A $\rho = 1$ Reversible Blinker.* The reversible blinker organisation contains processes that do not enable any other in the network, and have no edges pointing to their center-component (e.g., the processes in Figure 2.22a with an unspecified center-component). Such an organisation is degenerate in the sense that the interaction graph derived from it does not contain well-defined structures (Figure 2.22b). This arises because the components that would constitute the missing part of the boundary correspond to the problematic processes in the organisation, and thus the organisation fails to specify that those components should exist.

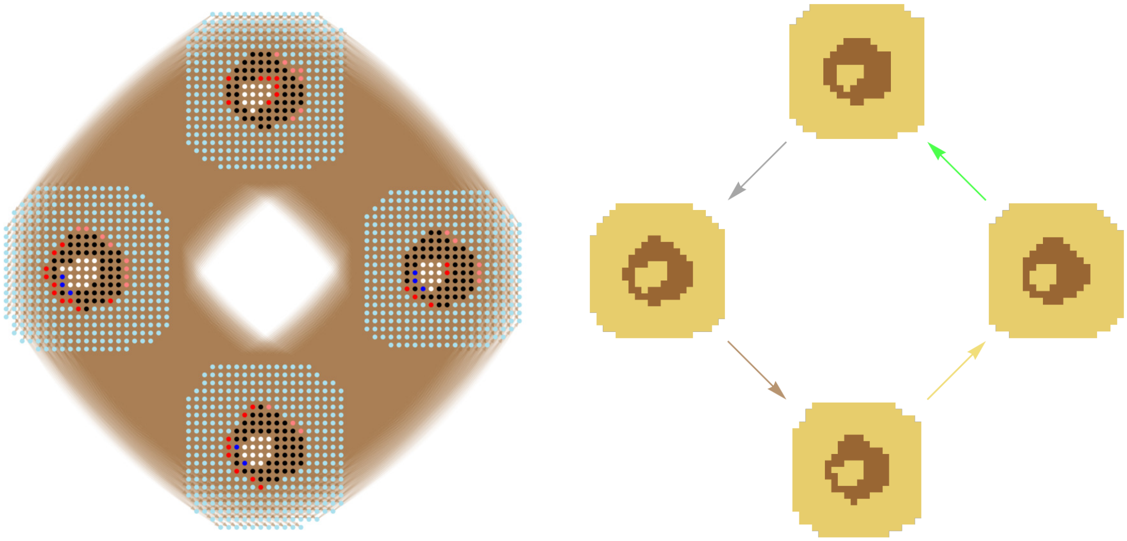
Unfortunately, the derivation procedure described above does not work in every case, as Example 10 shows. Specifically, there are degenerate cases where the organisation cannot be used to derive a proper interaction graph. The degeneracy arises due to the existence of processes that (i) enable no other processes in the organisation, and (ii) have no edges pointing to their center-component. More formally a vertex $i \in V$ corresponds to a degenerate process when there does not exist any edge with source vertex i nor any edge $(\cdot, i, k) \in E$ with $k = (\vec{0}, \cdot, \cdot)$. Reversible blinkers always have degenerate processes and are the only class of systems described in this chapter that do.

The significance of this depends on many factors. For one, we need to consider whether this very narrow class of degenerate cases should be taken to have the same theoretical weight as the more general case of bugs – are there any theoretical questions that blinkers allows us to answer that bugs do not? At least for the questions posed here, the answer to this is negative. Moreover, this degeneracy becomes *less* problematic, in a sense, for larger ρ blinkers, as the proportion of degenerate processes in the organisation (and thus proportion of missing components in the interaction graph) decreases with ρ. Moreover, it is not as though there is no way to recover the missing components: we can guarantee that a degenerate process is *always* a boundary process, and thus that its center-component is 0. Equivalently, one could "fill in" the missing components of the structure after the derivation to form a proper structure.

More fundamentally, though, this degeneracy is a consequence of the edge-decorations chosen for the definition of organisations and the special role that center-components play in the chemistry. We can define organisations using an alternative formulation by replacing the vertex set of arbitrary labels with properly defined processes. Then edge-decorations would only need to specify elements of $\mathscr{N}$, since the state of all reactants would be given explicitly by the vertices (implicitly, this is the formulation used in Beer's work on GoL: Beer, 2015, 2020b; Beer, McShaffrey, and Gaul, 2024). The reason I do not do this here is that such a representation would fail to scale into RealLife, as will be made clear in Chapter 3. Thus, the formalism presented here is a presented as part of a larger unified theoretical framework combining LtL and RealLife.

> *Example 11: A $\rho = 5$ $\bar{t} = 4$ Bug.* The interaction graph of the $\rho = 5$ $\bar{t} = 4$ bug in Example 5 (Figure 2.12) has 4 distinct spatial embeddings of the organisation, none of which are symmetries of each other. There are 4 edges and 4 non-destructive perturbation classes: $\mathcal{C} = \{\texttt{GRAY}, \texttt{BROWN}, \texttt{YELLOW}, \texttt{GREEN}\}$. Every class equivalent to a null-perturbation.

Example 11 shows the spatial embedding and interaction graph of a reversible organisation (Example 5). This interaction graph makes clear that every unique structure has a unique set of perturbation classes, even if that set only ever has one element. This is



(a) (b)

Figure 2.23: (a) The 4 spatial embeddings of the organisation in Example 5, with edges retained and labelled according to the interaction graph.(b) The interaction graph derived from the organisation in Example 5, with 4 perturbation classes labelled gray, brown, yellow, and green.

because perturbations can only be defined with respect to a structure – the geometry of a perturbation is contingent on the geometry of the structure it perturbs. Since every structure in this interaction graph has a unique geometry, there cannot be any shared perturbation classes. Moreover, perturbation classes must also be distinguished based on the internal configuration of a structure, even if geometries are identical. For instance, the glider's wedge and rocket structures have the same geometry, but differ in the particular configuration of components within that geometry (Example 8). It is necessary to distinguish perturbation classes here because the same perturbation may have different consequences – a perturbation applied to a rocket may destroy the glider organisation, but preserve it when applied to a wedge.

Just as there exist symmetric organisations in Example 5, there are symmetric interaction graphs here, where the same transformation can be applied to every structure of the graph, preserving the perturbation classes. There are a total of 8 such symmetric interaction graphs.

2.5. The Viability Constraint

Using this formalisation of autopoiesis and the cognitive domain, we can derive the set of environments in which an autopoietic system will persist. We call such a set of agent-environment configurations the *intrinsic viability constraint* of the system (Beer, McShaffrey, & Gaul, 2024). More formally, what we want is a set $\mathcal{V}$ of configurations that contain a realisation of the organisation. We further want to find a subset $\partial\mathcal{V} \subset \mathcal{V}$ in which every configuration destroys the organisation.

In order to properly formulate the viability constraint, we need to determine whether a given environment preserves the organisation of an autopoietic system. In general, this will require us to consider a structure and its (1,2)-environment – a set of the form $\mathcal{E}^2(\mathscr{S})$. The 1-environment is not sufficient when components of the system are produced by environmental processes, since those processes have reactants in the 2-environment.

There are two equivalent formulations of the intrinsic viability constraint. One specifies the conditions that universe state functions must satisfy with respect to the organisation, while the other specifies the conditions that *density functions* must satisfy. I call these the *state formulation* and the *density formulation*. (The former appeared in (Beer, McShaffrey, & Gaul, 2024), while the latter is novel.)

2.5.1. The State Formulation

The state formulation is quite simple. Given an autopoietic process dependency graph G and its derived interaction graph $\mathcal{I}_G = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$, $\mathcal{V}_G$ is the set of functions over $\mathcal{E}^2(\mathscr{S}_i)$ that contain some structure in $S_i \in \mathcal{S}$. $\partial\mathcal{V}_G$ is then the subset of $\mathcal{V}_G$ where the updated function does not contain any structure – the *boundary* of the viability constraint. All other sets can then be defined in terms of these two: the set of configurations that do not contain any realisation of the organisation is $\overline{\mathcal{V}_G}$, and the set of configurations that *preserve* the organisation is $\mathsf{int}(\mathcal{V}_G) := \mathcal{V}_G \backslash \partial\mathcal{V}_G$.

The following examples will illustrate this formulation more concretely.

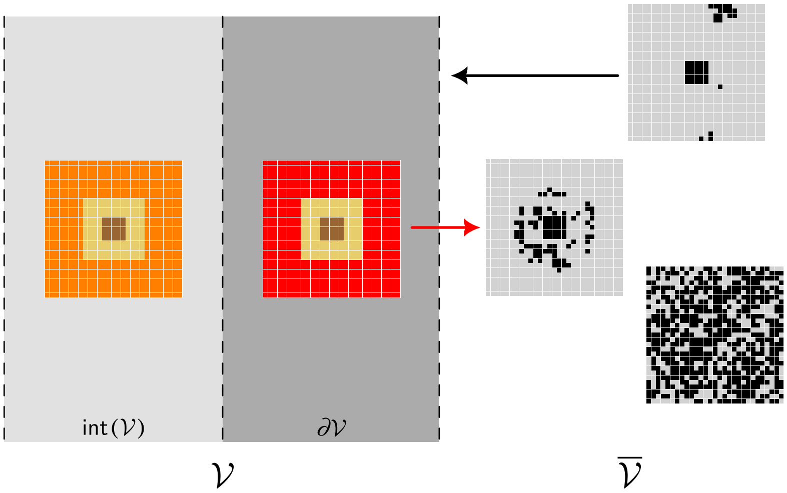
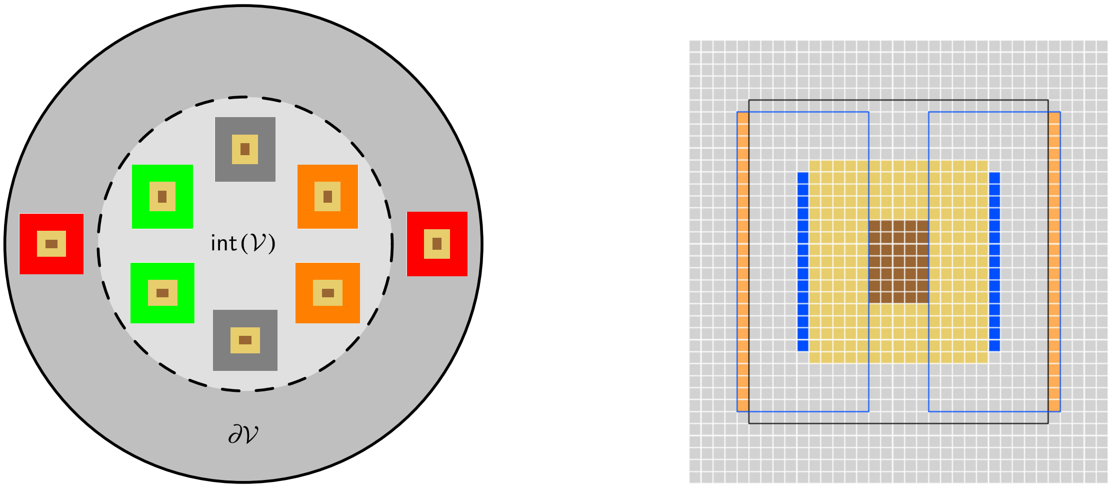


Figure 2.24: The intrinsic viability constraint of the $\rho = 4$ block. Orange indicates the class of non-destructive (1,2)-environments, and red indicates the class of destructive (1,2)-environments. The three configurations in $\overline{\mathcal{V}}$ are examples representing the kinds of configurations in that set. The top one will contain a block after one update, and thus has an arrow pointing to $\mathcal{V}$. The middle configuration is one that results from updating some member of $\partial\mathcal{V}$. The bottom one is neither the result of updating a member of $\mathcal{V}$, nor will update into $\mathcal{V}$.

> *Example 12: $\rho = 4$ Block.* Figure 2.24 shows a schematic illustration of the intrinsic viability constraint of a $\rho = 4$ block. There is 1 class of non-destructive (1,2)-environments in int($\mathcal{V}$) and 1 class of destructive environments in $\partial\mathcal{V}$. All members of $\partial\mathcal{V}$ invariably update to some member of $\overline{\mathcal{V}}$. Some configurations in $\overline{\mathcal{V}}$ will update to configurations in $\mathcal{V}$.

Still-lifes have the simplest possible viability constraints, with only two classes of (1,2)-environments in $\mathcal{V}$. However, this still captures most of the basic features common to all viability constraints. First, there are restrictions on what set a given configuration can transition to. Members of int($\mathcal{V}$) can only transition to other members of $\mathcal{V}$, since they by definition must preserve the organisation after updating. Similarly, members of $\partial\mathcal{V}$ always transition to members of $\overline{\mathcal{V}}$, by definition. More interesting are the possibilities for configurations in $\overline{\mathcal{V}}$: it is possible to transition to *any* of the sets.

Still-lifes are unique, however, in only needing the 1-environment to determine the viability constraint, since no components produced by the environment ever constitute the



(a)

(b)

Figure 2.25: (a) The intrinsic viability constraint of the $\rho = 5$ blinker organisation. There are 3 non-destructive classes, labelled green, gray, and orange. There is 1 destructive class, labelled red. (b) Components relevant to the gray class. In black is outlined the 1-environment of the brown and beige blinker configuration. The blue cells correspond to environmental processes that produce components that will constitute the blinker after updating. The blue outline contains all components in the neighbourhood of the blue cells. The orange cells are 2-environment components in the neighbourhood of the blue cells.

individual. Thus, the 2-environment is really unconstrained here, in that if any (1,2)-environment configuration is in a given set, all configurations that only differ by the 2-environment will also be in that set.

> *Example 13: $\rho = 5$ Blinker.* Figure 2.25a shows the intrinsic viability constraint of a $\rho = 5$ blinker. There are 3 classes of non-destructive (1,2)-environments and 1 class of destructive (1,2)-environments. The 2-environment is partially constrained in every class.

Example 13 describes a viability constraint with more than one non-destructive class of perturbations. Here the 1-environment is insufficient to fully-determine the viability constraint: in any given transition, there are always components that constitute the blinker that are produced by the environment. Since the neighbourhood of these processes intersects with 2-environment, their product can only be determined (in general) by accounting for the state of the 2-environment. However, this does not mean that *every* component of the

2-environment is constrained, but that only those in the neighbourhood of a process that produces components that constitute the blinker are constrained. For instance consider the GRAY class of (1,2)-environments (Figure 2.25b). Here, the components colored in blue will be transformed into the blinker boundary at the next step. Only the 1-environment and the orange components are relevant to this structural transition – all other components of the 2-environment can take any value.

It is worth noting that $\partial\mathcal{V}$ is not necessarily a boundary in the topological sense in the space of functions over $\mathcal{E}^2(\mathscr{S})$. For one, Each unique $\mathscr{S}$ would imply its own space of functions, which together need not have a connected union. More fundamentally, though, it is the *organisation* that determines what sets of function constitute $\mathcal{V}$, and this specification is independent of the topology in which these functions are considered.

2.5.2. The Density Formulation

While the state formulation is simple and captures what we want theoretically, it is practically limited. For instance, if one wants to find a perturbation in a given class, there is no obvious way to do this other than by searching the space of (1,2)-environments directly, or by solving a system of constraint equations (Beer, 2014). However, this is very expensive computationally. The density formulation simplifies this task by exchanging universe-state functions over $\mathcal{E}^2(\mathscr{S})$ for density functions over $\mathcal{E}(\mathscr{S})$. The main advantage here is that we can define a metric for how far a density function is from being a member of a given class. In fact, we can extend this more generally to determine how far a density function is from being in a set of classes (e.g., in $\mathsf{int}(\mathcal{V})$ or $\partial\mathcal{V}$).

The basic logic of the density formulation is grounded in results from Fourier analysis (Oppenheim & Schafer, 2010, pp. 49–50, 60). Specifically, a density function Ψ over some set $\mathscr{A} \subset \mathbb{Z}^2$ implies a *unique* $\mathbf{u} \in \mathcal{U}$ over $\mathcal{E}(\mathscr{A})$ such that $\Psi = \kappa * \mathbf{u}$. This function can be computed using the Discrete Time Fourier Transform (DTFT), notated here as $\mathcal{F}$:

$$\mathbf{u} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\Psi]}{\mathcal{F}[\kappa]}\right]. \tag{2.7}$$

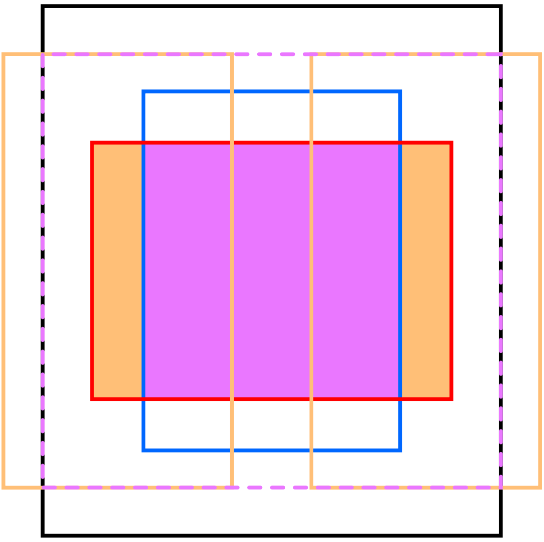


Figure 2.26: Schematic of a structural transition. The blue outline represents the initial configuration, and the red outline represents the configuration after updating. The black outline represents the 1-environment of the blue configuration. The violet region is the intersection of the red and blue configurations; the dotted violet line is the 1-environment of this region. The orange regions are the intersection of the red configuration with the 1-environment. The orange outlines are the 1-environments of the orange regions.

In order to derive $\mathcal{V}$, we again consider each edge $(S_i, S_j, \vec{v}) \in \mathcal{T}$ of the interaction graph $\mathcal{I}$. For each point in $\mathscr{S}_i \cap (\mathscr{S}_j + \vec{v})$ (the violet region in Figure 2.26), we can specify the bounds on the possible densities at those points such that $\xi[\mathbf{s}_i] = \sigma^{\vec{v}} \circ \mathbf{s}_j$, using the following equation:

$$\Gamma[\mathbf{a}, \mathbf{b}](x) := \begin{cases} [b_0, b_1] & \mathbf{a}(x) = 0 \wedge \mathbf{b}(x) = 1 \\ [0,1] \backslash [b_0, b_1] & \mathbf{a}(x) = 0 \wedge \mathbf{b}(x) = 0 \\ [s_0, s_1] & \mathbf{a}(x) = 1 \wedge \mathbf{b}(x) = 1 \\ [0,1] \backslash [s_0, s_1] & \mathbf{a}(x) = 1 \wedge \mathbf{b}(x) = 0. \end{cases} \tag{2.8}$$

Note that this function is only defined when both **a** and **b** are known – we need a *source* component and a *target* component. Thus, using $\Gamma[\mathbf{s}_i, \sigma^{\vec{v}} \circ \mathbf{s}_j]$, we get a set of density bounds for every point $x \in \mathscr{S}_i \cap (\mathscr{S}_j + \vec{v})$, which in turn implies a unique $\mathbf{u}$ over $\mathcal{E}(\mathscr{S}_i \cap (\mathscr{S}_j + \vec{v}))$

(the violet outline in Figure 2.26) for every Ψ that satisfies those bounds. However, this is not quite enough to determine the viability constraint in general, since it can only define at most the 1-environment by virtue of the density functions being constrained to some intersection with $\mathscr{S}_i$ (note how the dashed violet line is contained within the 1-environment in Figure 2.26).

The solution to this problem lies in applying Γ to the 1-environment states just derived. Before, we were missing source components for some of S_j (the two orange regions in Figure 2.26), but given a density function, these can be supplied. After applying Γ, we get a family of density function over $(\mathscr{S}_j + \vec{v}) \backslash \mathscr{S}_i$ (the orange regions). Now, if any of these density functions both satisfy Γ and are consistent with the density function used to derive the source components, we get a unique universe state function for all relevant components in the transition (union of the orange and violet outlines in Figure 2.26) – the transition $(S_i, S_j, \vec{v})$ is guaranteed to occur under this function, irrespective of the values of any other components in the (1,2)-environment.

The full viability constraint then becomes the set of valid density functions over each $\mathscr{S}_j + \vec{v}$ (union of the orange and violet regions) for every edge of the interaction graph. We can derive the state formulation of the viability constraint from this by replacing each density function with the set of (1,2)-environments that contain the components specified by that function.

To make this more concrete, I will derive a non-destructive perturbation for the organisation in Example 6. The interaction graph for this organisation is shown in Figure 2.27a. The edge in this graph has a translation vector $\vec{v} = \langle 0, 1 \rangle$, which means Γ is defined only over the brown and beige components in Figure 2.27b, resulting in the density bounds in Figure 2.27c. The density function in Figure 2.27d satisfies these bounds. Applying Equation 2.7, we set every component in the green region of Figure 2.27b to 0, which allows us to extend Γ over that region, assigning to each point $[0,1] \backslash [b_0, b_1]$ (both the source and target components are 0). These new bounds can be satisfied by setting the density at each point to 0. Finally, we apply Equation 2.7 again to get a state function with all relevant components in the (1,2)-environment set to 0, which can be easily verified as a viable non-destructive perturbation.

The most difficult part in using this method to derive perturbations is finding appropriate density functions. This is where the metric mentioned above becomes useful. We define

$$d_t[\Psi](x) := \min\left\{|\Psi(x) - y| : y \in \Gamma[\mathbf{s}_i, \sigma^{\vec{v}} \circ \mathbf{s}_j](x)\right\}, \tag{2.9}$$

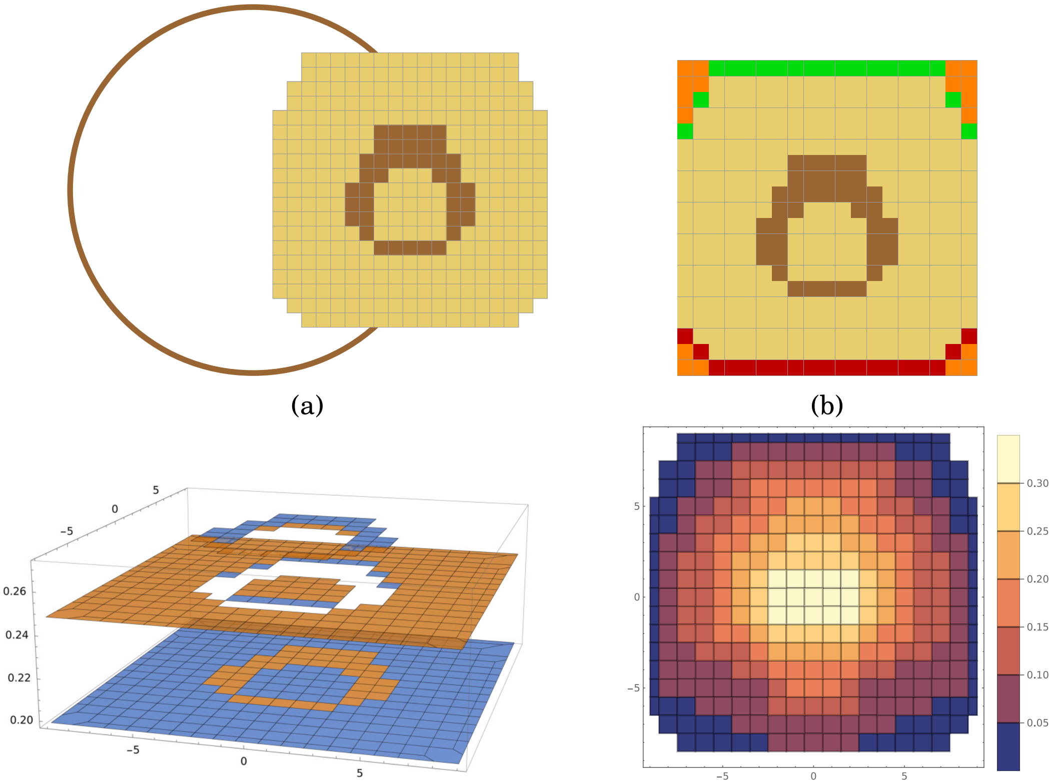
where $t = (S_i, S_j, \vec{v}) \in \mathcal{T}$ is an edge from an interaction graph, and Ψ is assumed to imply S_i under Equation 2.7. This gives the distance of how far a density function Ψ is from satisfying the transition t at a given point x. We can also extend this to be a metric over all transitions from a given structure S_i:

$$d_{\mathcal{V}}[S_i, \Psi](x) := \min\left\{d_t[\Psi](x) : t = (S_i, \cdot, \cdot) \in \mathcal{T}\right\}. \tag{2.10}$$

With this, the following procedure can be used to find appropriate density functions. First, begin with a structure S_i in a random (1,2)-environment and take the density function from this. Then manipulate this function through a process minimising the metric until a value of 0 is attained. Then for each possible transition satisfied by this density function, use Equation 2.7 to extend Γ and repeat the minimisation process. This should result in a valid density function, so long as the transition is possible at all (i.e., so long as Equation 2.7 returns a function in $\mathcal{U}$).

> *Example 14:* $\rho = 5$ $\bar{t} = 4$ *Bug.* Figure 2.28 shows the intrinsic viability constraint of the $\rho = 5$ $\bar{t} = 4$ bug from Example 5. There are 4 classes of non-destructive (1,2)-environments and 4 classes of destructive (1,2)-environments, where each of the 4 unique structures has exactly one class in each region of the viability constraint. The 2-environment is partially constrained in every class.

Returning to broader points about how viability constraints are defined in this formalism, consider Example 14. It is very clear to see how symmetries are treated: there are only



(a)

(b)

(c)

(d)

Figure 2.27: Deriving a perturbation using the density formulation. (a) The interaction graph derived from the organisation in Example 6. (b) Brown and beige cells correspond to processes in the organisation that produce components in the structure. Red cells correspond to processes in the organisation that do not produce components in the structure. Green cells correspond to environmental processes that produce components in the structure. Orange cells correspond to environmental processes that do not produce components in the structure. (c) The density bounds given by Γ. Blue bounds indicate that densities below the satisfy Γ, and orange bounds indicates that densities above satisfy Γ. (d) A density function that satisfies the density bounds in Figure 2.27c. Lighter color indicates higher density.

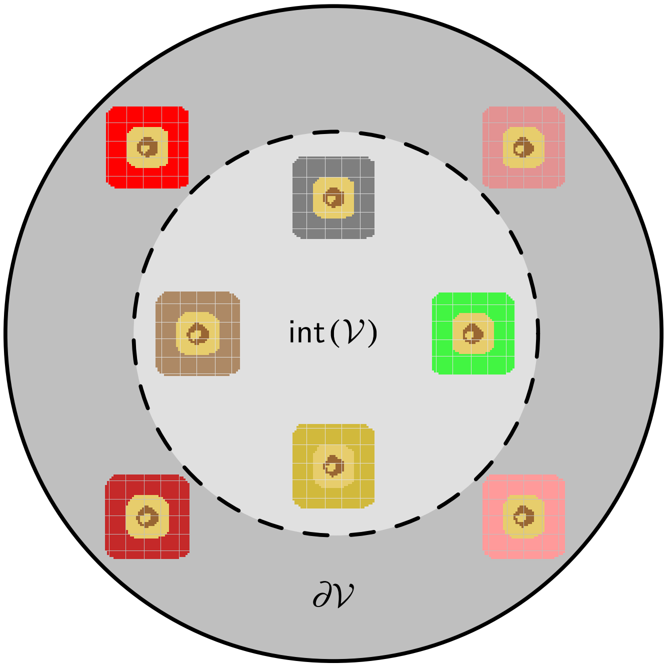


Figure 2.28: The intrinsic viability constraint in Example 14. The 4 classes of destructive (1,2)-environments in $\partial\mathcal{V}$ are colored in reds and pinks. The 4 classes of non-destructive environments in int($\mathcal{V}$) are colored in gray, brown, yellow, and green.

4 unique structures, but none of their symmetries. Thus, a configuration symmetric to one in the viability constraint does not realise the organisation. However, such a configuration may realise a symmetric organisation. This raises a reasonable concern about whether we should consider symmetric realisations as being of the same organisation. If we do, then any process dependency graph with distinct symmetries must properly be defined as a single disconnected graph, where each connected subgraph is one of the symmetric organisations. Similarly, the interaction graph would then also become a disconnected graph. Now, the viability constraint would contain all possible realisations, including symmetries. This formulation would add 56 new classes to Example 14, with 28 to int($\mathcal{V}$) and 28 to $\partial\mathcal{V}$.

There are many ways that this formalisation of viability differs from that presented in the literature thus far. Most formulations of viability essentially follow Ashby (1960, chap. 5), in which one considers some dynamical system and imposes constraints on its state space such that states within the constraint are, by definition, viable. The constrained dimensions are the *essential variables* of the system. The most obvious difference between this and the formulation above is where the constraint comes from: in one case, the constraints are given by the observer independent of the system considered, where as in the other, the constraints are derived directly from the system itself. This is why I have used the term *intrinsic* viability constraint, differentiating this from the *extrinsic* constraints *a la* Ashby (Ashby, 1960; Beer, McShaffrey, & Gaul, 2024). A particularly salient consequence of the intrinsic-extrinsic dichotomy is how, in the extrinsic case, the system still exists in some sense, even if its state is outside the viability region. This differs from intrinsic constraints, in which it does not make sense to talk about the state of an individual when it is not realised.

Another basic difference is that intrinsic constraints restrict on both the space of structures *and* structural transitions, whereas extrinsic constraints only restrict the states (structures) of a system. Relatedly, the environment need not be considered in the extrinsic case unless introduced explicitly by the observer (for instance, by adding a behavioural mechanism effecting the essential variables (Barandiaran & Egbert, 2014; McShaffrey & Beer, 2023), or by making the system nonautonomous (Aubin, 1991)). Moreover, even in the case of extrinsic constraints on nonautonomous dynamical systems, one does not exhaust all possible environments as was done above, and so the observer still plays a role in deciding what environmental variables or behavioural dynamics are relevant.

It is also interesting to consider how the concept of essential variables maps to the intrinsic case. If we understand an essential variable to be a variable such that changes in its state can change the region of the viability constraint the system as a whole is in, then the essential variables implied by an intrinsic constraint include not only the components of the structure, but also environmental components. For instance, consider Example 14 where, for any given structure, the system could be in either $\mathsf{int}(\mathcal{V})$ or $\partial\mathcal{V}$, depending on the state of the (1,2)-environment. Moreover, what set of essential variables are relevant at a given time will change depending on both the structure and transition considered; in Example 13, the set of relevant (1,2)-environment components differ between each class. Even more interestingly, some of the essential variables may be in the 2-environment of a structure, and thus the information available to the system (its internal state and the 1-environment) becomes insufficient to determine what region of the viability constraint it is in.

Finally, it is worth commenting on the sense in which the intrinsic viability constraint is derived from the organisation. Most obviously, the derivation requires us to consider (1,2)-environments, which is information necessarily external to the organisation. We also need to supply an LtL rule in order to determine the constraint. Regarding the first, it should be clear that the space of all (1,2)-environments is not exactly new information as much as it is the largest space of possibilities in which we can consider a system. It should also be evident that all restrictions on this space are derived from the organisation, since the organisation alone specifies all viable structures and transitions (Section 2.4). Regarding the requirement of an LtL rule, this depends on whether we interpret rules as fixed parameters of the universe, or biologically relevant variables. But even in the latter case, we can, under specific circumstances, derive the set of rules that support an organisation (this is done in the next section). Thus, one could also consider the intrinsic viability constraint to be a set of viability regions $\mathcal{V} = \mathsf{int}(\mathcal{V}) \cup \partial\mathcal{V}$, one for each rule. In any case, an intrinsic viability constraint is still derived purely from the organisation, as all restrictions on the full space of possibilities arise from information implicit in the process dependency graph.

2.6. Rule Derivations

This section describes a procedure by which one can derive the set of LtL rules that support an organisation, in the sense that every transition in the interaction graph is possible when all relevant components of the (1,2)-environment are 0-components. This means that the following will only apply to *reversible* organisations (or reversible subgraphs of a larger a organisation.)

The procedure for deriving the set of rules is as follows. We define an 8-tuple of parameters $(b_0^L, b_0^U, b_1^L, b_1^U, s_0^L, s_1^U)$, where

$$b_0^L < b_0 \leq b_0^U \tag{2.11}$$

$$b_1^L \leq b_1 < b_1^U \tag{2.12}$$

$$s_0^L < s_0 \leq s_0^U \tag{2.13}$$

$$s_1^L \leq s_1 < s_1^U \tag{2.14}$$

$$0 < s_0 \leq b_0 \leq b_1 \leq s_1 \leq 1. \tag{2.15}$$

Thus, we are trying to derive bounds on each LtL rule parameter such that any $(\rho, b_0, b_1, s_0, s_1)$ that satisfies the bounds is guaranteed to support the organisation. Note also that Equation 2.15 is a modification of Equation 2.3, where now we require $b_0 > 0$. This is to ensure that a universe of all 0-components is stable, consistent with our assumption about 0-components in the (1,2)-environment.

The first step to finding values for these parameters is to partition the processes by their type: production ($\mathcal{X}^b$), 0-maintenance ($\mathcal{X}^{\neg b}$), 1-maintenance ($\mathcal{X}^s$), and destruction ($\mathcal{X}^{\neg s}$). Given our assumption about the 1-environment, this means that partial processes are now treated as fully determined processes, with all unspecified components set to 0. Let $G = (V, E, \mathfrak{s}, \mathfrak{t})$ be a process dependency graph; then

$$\mathcal{X}^b := \left\{\psi_i : \mathbf{p}_i(\vec{0}) = 0 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 1)\right\} \tag{2.16}$$

$$\mathcal{X}^{\neg b} := \left\{\psi_i : \mathbf{p}_i(\vec{0}) = 0 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 0)\right\} \tag{2.17}$$

$$\mathcal{X}^s := \left\{\psi_i : \mathbf{p}_i(\vec{0}) = 1 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 1)\right\} \tag{2.18}$$

$$\mathcal{X}^{\neg s} := \left\{\psi_i : \mathbf{p}_i(\vec{0}) = 1 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 0)\right\}. \tag{2.19}$$

Here k_{c_1} is the component produced by the process (c_1 in $k = (\vec{r}, c_0, c_1)$), and $\psi_i = (\phi_i, \mathbf{p}_i)$ is the process associated with the label $i \in V$. Note that these sets will exclude any processes that do not enable anything, since the products of such processes are always irrelevant.

Now we can compute the 8 parameters:

$$b_0^U := \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} \tag{2.20}$$

$$b_0^L := \max\left\{\Psi(\vec{0}) : \Psi(\vec{0}) < b_0^U,\ \psi \in \mathcal{X}^{\neg b}\right\} \tag{2.21}$$

$$b_1^L := \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} \tag{2.22}$$

$$b_1^U := \min\left\{\Psi(\vec{0}) : \Psi(\vec{0}) > b_1^L,\ \psi \in \mathcal{X}^{\neg b}\right\} \tag{2.23}$$

$$s_0^U := \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} \tag{2.24}$$

$$s_0^L := \max\left\{\Psi(\vec{0}) : \Psi(\vec{0}) < s_0^U,\ \psi \in \mathcal{X}^{\neg s}\right\} \tag{2.25}$$

$$s_1^L := \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} \tag{2.26}$$

$$s_1^U := \min\left\{\Psi(\vec{0}) : \Psi(\vec{0}) > s_1^L,\ \psi \in \mathcal{X}^{\neg s}\right\}, \tag{2.27}$$

where $\Psi = \kappa * \mathbf{p}$ is the (lower bound) density of the process ψ. The basic idea in the definition of these parameters are to find the processes in the network with the highest or lowest density in any set $\mathcal{X}$. For example the inclusive upper bound for b_0, b_0^U, is defined as the smallest density of a production process. If b_0 were any larger, then that process would become 0-maintenance, and thus the organisation would no longer be realised.

However, some qualifications need to be made about these definitions, since some may of the values not be defined when any of the process sets $\mathcal{X}$ are empty (i.e., when not all 4 process classes are present in the organisation). If any of the parameters are undefined, they can be ignored in the inequalities. However, when the parameter is needed to compute another (e.g., b_0^U in b_0^L), more care is needed. Specifically, if $\mathcal{X}^b$ is empty – i.e., b_0^U and b_1^L are undefined – b_0^U should be ignored in the definition of b_0^L, and b_1^U should be ignored in the inequality 2.12. Otherwise, if a parameter does not exist, it can be ignored in any definition that it appears in.

The logic of and motivation for this procedure is described in detail in Section A.1.4, but it is basically grounded in three basic assumptions: (i) there is a 0-maintenance process with $\Psi < b_0$; (ii) if there is a 0-maintenance process with $\Psi > b_1$, $\mathcal{X}^b$ must be non-empty; and (iii) $\mathcal{X}^s$ and $\mathcal{X}^{\neg b}$ are always non-empty (i.e., there are always 1-maintenance and 0-

maintenance processes). These assumptions limit the power of the method, as it may not be exhaustive when any of the sets $\mathcal{X}$ do not exist. For instance, a block can be supported by a rule in which all 0-maintenance processes have $\Psi > b_1$, violating assumption (i). However, every rule derived by this procedure is guaranteed to support the organisation when every relevant component in the (1,2)-environment is 0. Moreover, in the case where every $\mathcal{X}$ does exist, the derived rules should include *all* and *only* those rules that support the organisation under those same environmental conditions.

Example 15: $\rho = 1$ *Block.* Using the organisation in Example 1, we get 12 processes in $\mathcal{X}^{\neg b}$, 4 processes in $\mathcal{X}^{s}$, and all other sets empty. This gives

$$\begin{aligned} b_0^L &= \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^{\neg b}\right\} = \frac{2}{9} \\ s_1^L &= \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^{s}\right\} = \frac{4}{9} \\ s_0^U &= \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^{s}\right\} = \frac{4}{9}. \end{aligned}$$

Therefore, every rule $(1, b_0, b_1, s_0, s_1)$ that satisfies

$$\begin{gathered} b_0 > \frac{2}{9} \\ s_0 \leq \frac{4}{9} \leq s_1 \\ 0 < s_0 \leq b_0 \leq b_1 \leq s_1 \leq 1 \end{gathered}$$

will support the $\rho = 1$ block when every component of the 1-environment is 0.

As mentioned above, there are rules that support a $\rho = 1$ block but are not accounted for by Example 15. For instance, any rule with $s_0 = b_0 = b_1 = 0$ and $s_1 \geq \frac{4}{9}$ will also support the block, even when the 1-environment contains 1-components.

Example 16: Reversible Gliders. Consider a reversible subgraph of the glider organisation (Example 7) corresponding to any one of the 4-cycles in Figure 2.19c with black and grey edges. From this, we get 8 processes in $\mathcal{X}^{b}$, 60 in $\mathcal{X}^{\neg b}$, 12 in $\mathcal{X}^s$, and 8 in $\mathcal{X}^{\neg s}$. This gives us

$$\begin{aligned}
b_0^U &= \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} = \frac{3}{9} \\
b_0^L &= \max\left\{\Psi(\vec{0}) < b_0^U : \psi \in \mathcal{X}^b\right\} = \frac{2}{9} \\
b_1^L &= \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} = \frac{3}{9} \\
b_1^U &= \min\left\{\Psi(\vec{0}) > \frac{3}{9} : \psi \in \mathcal{X}^{\neg b}\right\} = \frac{4}{9} \\
s_0^U &= \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} = \frac{3}{9} \\
s_0^L &= \max\left\{\Psi(\vec{0}) < \frac{4}{9} : \psi \in \mathcal{X}^{\neg s}\right\} = \frac{2}{9} \\
s_1^L &= \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} = \frac{4}{9} \\
s_1^U &= \min\left\{\Psi(\vec{0}) > \frac{4}{9} : \psi \in \mathcal{X}^{\neg s}\right\} = \frac{5}{9}.
\end{aligned}$$

Plugging these values into the inequalities, we get exactly one rule:

$$\left(\rho = 1, b_0 = \frac{3}{9}, b_1 = \frac{3}{9}, s_0 = \frac{3}{9}, s_1 = \frac{4}{9}\right).$$

Therefore, gliders can only exist under the GoL rule.

It can be easily verified that the GoL rule also satisfies the constraints derived in Example 15.

The significance of these derivations depends on how we interpret LtL rule parameters physically. If we consider them simply as static parameters defining the unchanging physical laws of an LtL universe, then their derivation from the organisation is less biologically interesting – no *biological* problem is solved if we derive quantum mechanics from the organisation of a cell. However, if we interpret LtL rules as a more general physical parameter, then there is something of interest, especially regarding viability. For instance, the parameters could be interpreted as something akin to temperature (modulating the sensitivity of reactions to changes in the proportion of reactants), and thus the rule derivation would be akin to deriving the range of temperatures in which a cell could survive. If the derivation is restricted to subgraphs of the organisation, then this would be akin to finding the range of temperatures in which a cell exhibits a certain behaviour (e.g., sporulation, virulence, etc.).

This derivation procedure also offers an alternative to Evans' (2003) approach to finding rules for morphologically similar bugs at increasing ρ. For example, given a translation-invariant bug at some ρ, Evans' method for finding a similar bug at $\overline{\rho}$ involved rescaling the initial bug by $\rho/\overline{\rho}$ and searching around a scaled rule (using a formula quadratic in ρ; Evans, 2003, p. 51) until the dynamics settle into a bug with the desired properties. Unfortunately, this method involves a fair amount of experimentation. Using the rule derivation, however, a more direct method is possible. The first step is the same: scale an initial bug by $\rho/\overline{\rho}$. The next step is to "smooth out" this larger configuration by, for example, applying a convolution (not necessarily the same as in Equation 2.1). Then one could specify a translation vector $\vec{d}$ to get an interaction graph, and then derive a process dependency graph from this. Finally, the rule derivation would give a set of rules that supports the scaled bug. This method has advantages in both not requiring experimentation and providing more than just one rule.

Whether it is possible to similarly derive rules for non-reversible organisations is at this point unclear. The main difficulty is that processes could no longer be treated in isolation, and thus the spatial embeddings of the organisation would need to be considered in order to determine what process densities could be simultaneously realised. This forces us to restrict both the space of rules and the space of (1,2)-environments simultaneously. An intermediate possibility would be to assume that every structure in an interaction graph belongs to some reversible subgraph from which a set of rules can be derived. Then for each such rule, one could search for perturbations satisfying the remaining edges of the interaction graph (using the method described in Section 2.5). Then every rule where a perturbation exists for every edge supports the full non-reversible organisation.

***

The theory of autopoiesis presented in this chapter applies to any LtL universe of any ρ. However, showing that there exists autopoietic systems for any ρ is non-trivial. Evans (2003) provides rigorous results that prove that $(\rho+1)\times(\rho+1)$ blocks and $\rho\times(\rho+2)$ blinkers exist for all finite $\rho > 0$. Unfortunately, no such results exist for bugs with stomachs, though there are empirical results suggesting that translation-invariant bugs can be scaled arbitrarily, as mentioned above in Section 2.1 (Evans, 2003).

Given that finding a formula for rules supporting translation-invariant bugs is so difficult, an alternative approach could be to characterise limiting sequences of process dependency graphs and show that the limit extends indefinitely. Then, for any graph in the sequence, a supporting rule can be derived. Fortunately, there are rigorous results that show that there is a continuum limit of LtL, called *RealLife* (Pivato, 2007). Moreover, the existence of still-lifes in this limit has been proven. The next chapter develops a theory of autopoiesis in RealLife, and defines the convergence of sequences of LtL organisations to this limit.

Chapter 3

Autopoiesis in RealLife

This chapter develops a formal framework for analysing autopoiesis in RealLife Euclidean automata (EA). First, I will define RealLife and discuss some of the rigorous results relating it to LtL. Then I will extend the interpretation of autopoiesis presented in Chapter 2 to RealLife, before making this interpretation mathematically explicit in Section 3.2. Section 3.3 will show how autopoiesis in LtL can be treated as an approximation to RealLife, or equivalently, how the theory presented in this chapter is the continuum limit of the LtL theory. The remaining sections present the rest of the RealLife theory: the cognitive domain (Section 3.4), the intrinsic viability constraint (Section 3.5), and rule derivations (Section 3.6). A rigorous statement of the theory presented in this chapter can be found in Appendix A.2.

3.1. RealLife

RealLife is a family of D-dimensional *Euclidean automata* (Pivato, 2007) based on LtL (Evans, 2001). The main difference is that $\mathbb{Z}^D$ is exchanged for $\mathbb{R}^D$. Thus, keeping the same $\mathscr{M} = \{0,1\}$, the state of a RealLife universe is given by a function $\mathbf{u} : \mathbb{R}^D \to \mathscr{M}$, assigning a value to each point in D-dimensional space; again, I notate the space of these functions $\mathcal{U}$. The RealLife neighbourhood is specified by a choice of p-norm and a finite radius $\rho \in \mathbb{R}^+$: $\mathscr{N} := \left\{x : \|x\|_p \leq \rho,\, x \in \mathbb{R}^D\right\}$. I will usually assume $p = \infty$, except when stated otherwise. I will always assume $D = 2$.

The density function Ψ is defined analogously to LtL:

$$\kappa(y) := \lambda\,[\mathscr{N}]^{-1}\,\mathbb{1}_{\mathscr{N}}[y] \tag{3.1}$$

$$\Psi(x) := \kappa * \mathbf{u}(x) = \lambda[\mathscr{N}]^{-1} \int\limits_{y \in \mathscr{N}} \mathbf{u}(x-y)\,\lambda[dy], \tag{3.2}$$

where λ is the D-dimensional Lebesgue measure on $\mathbb{R}^D$ (a function giving the volume of a region.) A RealLife EA is then defined as on operator $\xi : \mathcal{U} \to \mathcal{U}$:

$$\xi[\mathbf{u}](x) := \mathbf{u}(x)\mathbb{1}_{[s_0,s_1]}(\Psi(x)) + (1-\mathbf{u}(x))\,\mathbb{1}_{[b_0,b_1]}(\Psi(x))\,. \tag{3.3}$$

The rule parameters $(\rho, b_0, b_1, s_0, s_1)$ differ slightly from LtL. Most obviously ρ is now a real number, as stated above. Moreover, the inequality that the parameters must satisfy now includes $0 < s_0$ and $b_0 < b_1$ (instead of $0 \leq s_0$ and $b_0 \leq b_1$):

$$0 < s_0 \leq b_0 < b_1 \leq s_1 \leq 1. \tag{3.4}$$

RealLife EAs are *continuous space, discrete time universes*. That is to say, ξ is not a differential equation – the dynamics can only evolve in discrete steps. Thus, patterns in RealLife are still described by finite sets of functions, and we can directly carry over the classes of emergent patterns in LtL: a still-life is a function $\mathbf{u} = \xi[\mathbf{u}]$; an oscillator $\mathbf{u} = \xi^{\bar{t}}[\mathbf{u}]$, for some $\bar{t} \in \mathbb{N}$; and a bug $\sigma^{\vec{d}} \circ \mathbf{u} = \xi^{\bar{t}}[\mathbf{u}]$ for some $\bar{t} \in \mathbb{N}$ and $\vec{d} \in \mathbb{R}^D$.

Pivato (2007, Theorem 2.1) proved that RealLife EA is the limit of LtL CAs as $\rho \to \infty$. This naturally raises the question whether still-lifes, oscillators, and bugs also exist in RealLife. Fortunately, Pivato (2007, Proposition 3.4) has provided rigorous results that show that a generalisation of the block exists (however, no proofs yet exist for the other classes). Thus, it is natural to try and generalise the theory presented in Chapter 2 to RealLife. This would also show that this interpretation of autopoiesis has power beyond CAs.

Note that there are different ways to interpret the formal relation between LtL and RealLife, and what this means for the autopoietic theory. At the end of Chapter 2, RealLife was introduced as the infinite limit of LtL – practically, RealLife as demonstrating indefinite *discrete* scaling, and (potentially) as tool to approximate very large LtL CAs. However, we can invert this relation: LtL as a discrete approximation to RealLife. These are equivalent interpretations (in fact, Pivato (2007, Theorem 2.1) proves RealLife is the limit of LtL by showing that LtL can approximate RealLife arbitrarily well). Practically, we can treat either as an approximation of the other depending on what we are interested in and what is easier to compute.

Before laying out the details of the theory in RealLife, it will be useful to map our interpretation of autopoiesis more broadly. First, the chemistry remains largely the same (Figure 3.1a): a process is a pair $\psi = (\phi \subseteq \mathscr{N}, \mathbf{p} : \phi \to \mathscr{M})$, with product $\xi[\mathbf{p}](\vec{0})$. The main difference here is that there are now infinitely many reactants in ϕ. Process dependency also extends directly: ψ_i at x_i and t_1 enables ψ_j at x_j and t_2 if $\|x_i - x_j\|_p \leq \rho$ and $t_2 = t_1 + 1$. The reactant that ψ_i produces in ψ_j is $x_i - x_j \in \phi$, whose state is the product of ψ_i.

Structures similarly extend from LtL in a straightforward manner. All that changes is that $(\mathscr{S}, \mathbf{s})$ now has $\mathscr{S} \subset \mathbb{R}^2$. The notion of n-environments also applies: the set of points in $\mathscr{S}$ and its 1-environment, $\mathcal{E}(\mathscr{S})$, is the set of points in the neighbourhood of some $x \in \mathscr{S}$;

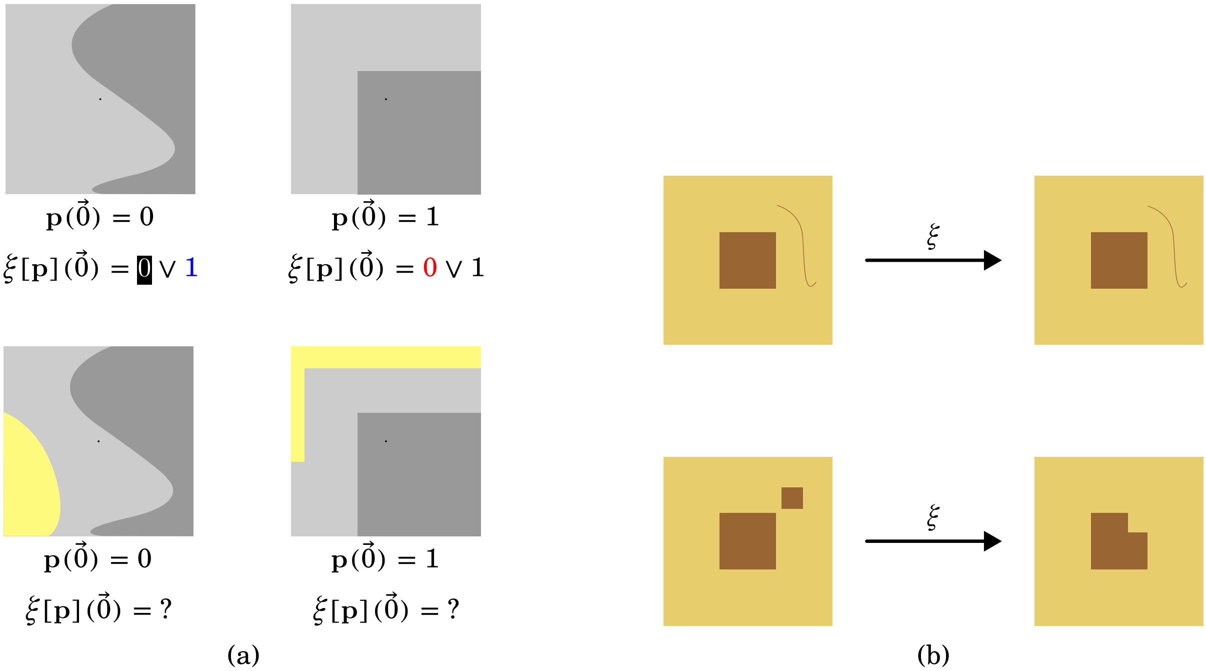

$\mathbf{p}(\vec{0}) = 0$
$\xi[\mathbf{p}](\vec{0}) = 0 \vee 1$
$\mathbf{p}(\vec{0}) = 1$
$\xi[\mathbf{p}](\vec{0}) = 0 \vee 1$
$\mathbf{p}(\vec{0}) = 0$
$\xi[\mathbf{p}](\vec{0}) = ?$
$\mathbf{p}(\vec{0}) = 1$
$\xi[\mathbf{p}](\vec{0}) = ?$
(a)
ξ
ξ
(b)

Figure 3.1: (a) Process classes in RealLife. The top right process is a either a 0-maintenance or a production process depending on the rule supplied, since it has a center-component of 0. The process to the right of that has a center-component of 1 and is either a destruction or a 1-maintenance process. The bottom two processes are partial versions of the above two, with $\phi \subset \mathscr{N}$. The product of these process cannot be determined. (b) Dynamics of structures in RealLife. On top is a structure that only differs from the block by a set of 0 area. This structure updates to a function that differs at most by that area. On the bottom is a block structure with a set of 1-components inserted into its boundary. After updating, this will result in changes to the structure beyond the modified region.

the n-environment is attained by recursing this function: $\mathcal{E}^n$. Again, the main difference here is that there are now infinitely many components in the structure.

These infinities introduce a new complication not present in LtL: two structures (or two processes) can differ on some subset of their components, but be “effectively the same.” Specifically, if two functions $\mathbf{a}, \mathbf{b} \in \mathcal{U}$ differ on some set with an area of zero ($\lambda[\mathscr{A}] = 0$), then $\xi[\mathbf{a}]$ and $\xi[\mathbf{b}]$ will similarly only differ at most by that set. For example, if one modifies a single component of a structure, the global dynamics of that structure will be unchanged; if a set of components with non-zero area is modified, however, the structure may be destroyed just as in LtL (Figure 3.1b). Intuitively, then, two structures that differ only by a set of non-zero area should be considered the same. This will simplify aspects of the theory developed below by allowing us to describe structures by a single functions, ignoring all other functions that are *almost everywhere equal* (differ by a set of zero area).

My interpretation of what constitutes an emergent individual will also remain the same as in LtL. That is, an individual is defined by its organisation such that every realisation of the organisation is described by a structure $(\mathscr{S}, \mathbf{s})$ with a ρ-wide boundary of 0-components (equivalently, we can state that $\mathscr{S}$ is equal to its 1-components and their 1-environment). The dynamics of these structures must also be such that if a process in $S_i = (\mathscr{S}_i, \mathbf{s}_i)$ enables some process in S_j, then every process in S_j is enabled by a process in S_i: S_j must be contained within the light cone of S_i $((\mathscr{S}_j + v) \subseteq \mathcal{E}(\mathscr{S}_i))$. I will also only be considering organisations whose set of structures is finite. In principle, organisations with infinite structure can exist in RealLife, but many aspects of the theory developed here fail in that case. However, any organisation scaled from LtL will have a finite structure set, so this does not impact the theoretical interest of the theory as a limit of LtL.

3.2. Organisation

This section will generalise process dependency graphs to describe organisations in RealLife. But first, it will be productive to recall the desiderata we want the representation of an autopoietic organisation to satisfy. Most obviously, it will need to describe dependency relations between processes, and specify those processes. It should also be invariant with respect to how processes are arranged – the representation of a process should tell us nothing about its spatial embedding, except via dependencies. Finally, we should be able to derive objects describing the cognitive domain and the intrinsic viability constraint.

The appropriate object to describe autopoietic organisations in RealLife is a *graphon* (graph function; Lovász, 2012). I will first give the definition more formally and then give an intuitive description of how it represents an autopoietic network and satisfies the above desiderata.

Let $K := \mathscr{N} \times \mathscr{M} \times \mathscr{M}$ be a set with elements $k = (k_r, k_{c_i}, k_{c_j})$, and let K_0 be K with a special 0 element (it does not have to be 0, just an arbitrary element not in K). Let $\mathbb{P}(K_0)$ denote the space of probability measures on K_0. Then a *process dependency graphon* is a function $\omega : \Omega \times \Omega \to \mathbb{P}(K_0)$, where Ω is an arbitrary set.

We interpret this definition as follows. First, Ω is a set of process labels. $\omega(x,y) = \mu$ then assigns an edge decoration to every pair of processes x and y. The edge decoration μ is unfortunately a probability measure, which is not very useful. However, if we assume some bijection $g : k \mapsto \mu$, and that ω only gives measures in the image of g, then we can define a new function $W(x,y) := g^{-1} \circ \omega(x,y)$ that gives edge decorations in K_0. These are the same edge decorations as in LtL, except we now have the special 0 element which we interpret as indicating no dependency relation. In other words, $W(x,y) = k = (r, c_0, c_1)$ gives the dependency relation for every pair of processes x and y, where k_r is the component in y produced by x, k_{c_1} is the state of that component, and k_{c_0} is the state of the center-component of x (to simplify notation, I will denote these by $W_r(x,y)$, $W_{c_1}(x,y)$, and $W_{c_0}(x,y)$, respectively). The reason for using probability measures will become clear in Section 3.3 when we need to define the convergence of graphs to graphons. For the autopoietic theory, though, we can restrict our attention exclusively to W.

Just as in LtL, process functions are implicit but recoverable. Specifically, given some $x \in \Omega$, we get $\phi_x = \{W_r(\cdot, x)\}$ as the set of all k_r where $W(\cdot, x) = k$, and $\mathbf{p}_x(z) = W_{c_1}\left(\overleftarrow{W}^z(x), x\right)$, where $\overleftarrow{W}^z(x)$ gives the set of processes y that enables x with $W(y,x) = (z, \cdot, \cdot)$.

We can define the *closure* of an organisation W by relabelling all edges pointing to processes that do not enable anything:

$$W_{closure}(x,y) := \begin{cases} W(x,y) & \exists y' \text{ s.t. } W(y,y') \in K \\ 0 & \end{cases}. \tag{3.5}$$

The notion of *reversible organisations* also generalises. An organisation is reversible when every reactant of every process is produced by exactly one process. More technically, we sat that W is reversible if, for every $y \in \Omega$ and $k = (r, \cdot, \cdot) \in K$, there is at most one x such that $W(x,y) = k$.

This definition of process dependency graphons allows us to have infinitely many processes and edges, as Ω need not be a finite set. Moreover, we need not assume any specific topology on Ω: processes can be randomly rearranged, so long as ω is changed accordingly. In other words, the specific representation of an organisation is not, strictly speaking, unique – two representations ω_i and ω_j are equivalent if there is some map $f : \Omega_i \to \Omega_j$ such that $\omega_i(x,y) = \omega(f(x), f(y))$. Similarly, the choice of g is arbitrary. Therefore, any result of the theory obtained for one representation should still hold for all equivalent representations.

One last point bears mentioning here. *This formalisation does not generalise multigraph organisations*. That is, $W(x,y)$ is unique for every ordered pair (x,y). In principle, one could represent multigraphs that have an upper bound n on the number of parallel edges by using K_0^n (Lovász, 2012, p. 324). However, this would significantly complicate the theory and set an unnecessarily arbitrary limit on what organisations can be represented. It is also unclear whether graphons with unbounded parallel edges could be used, but again, this would make things quite complicated.

We can now begin looking at examples of autopoietic organisations in RealLife, noting some of their significant features and the practical issues that arise in finding an appropriate representation.

Example 1: The Block. Let $\Omega_{Block} := \odot_{\frac{3\rho}{2}}\left(\vec{0}\right)$ where

$$\bigodot_{r}(x) := \left\{x : \|x - y\|_p \leq r\right\}.$$

Then the *block organisation* is defined

$$W_{Block}(x,y) := \begin{cases} \left(x - y, \mathbb{1}_{\odot_{\rho/2}(\vec{0})}(x), \mathbb{1}_{\odot_{\rho/2}(\vec{0})}(x)\right) & \|x - y\|_p \leq \rho \\ 0 & \end{cases}. \qquad (3.6)$$

W_{Block} is reversible and equivalent to its closure.

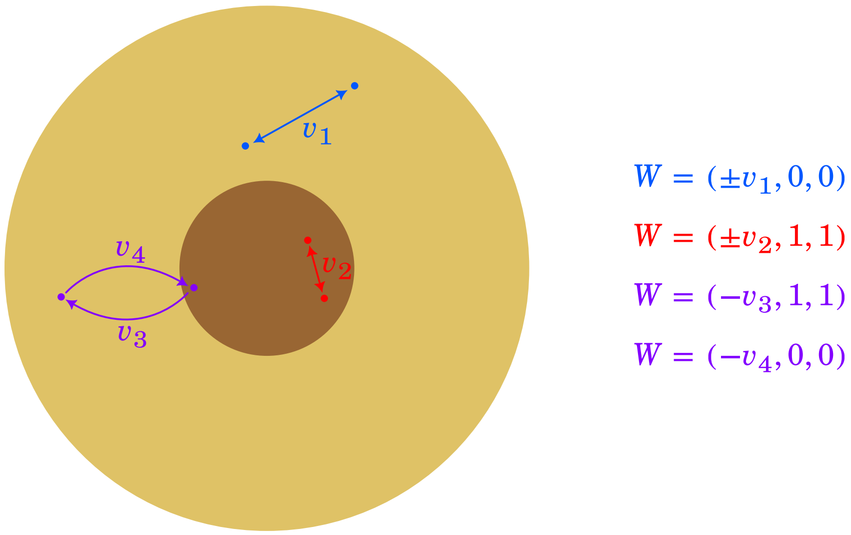


Figure 3.2: Schematic of the $p = 2$ block organisation. Beige and brown points are 0- and 1-components, respectively. In red is a pair of edges between 1-maintenance processes, with the equation describing this edge in the corresponding color on the right. In blue is a pair of edges between 0-maintenance processes. In purple is a pair of edges between a 1-maintenance and a 0-maintenance process. Edges between processes of different types cannot be reciprocated by inverting the edge.

The block will be our most lucrative example, as it is the simplest object we will consider and is the only one proven to exist in RealLife, as mentioned above. Example 1 defines the organisation of the block mathematically (Figure 3.2 provides a schematic illustration). This representation is quite simple and takes advantage of our knowledge of the block's morphology. Ω_{Block} is simply the set of points that constitute a block centered at the origin: the process $x \in \Omega_{Block}$ is located at $\sigma^v(x)$ for a block centered at v. Recall that, so long as the theory holds for all equivalent representations, we can use any such representation we like – we can use knowledge of the structure so long as the theory does not depend on that knowledge.

The two cases in W_{Block} correspond to whether a dependency relation exists between x and y (whether x is in the neighbourhood of y). The result in the first case is an edge decoration in K. $x - y \in \mathcal{N}$ is the vector from y to x which is equivalent to the reactant in ϕ_y that x produces ($-v$ in Figure 3.2). The two indicator functions give the center-component of x. Again, since Ω_{Block} is just the components of the block's structure, we

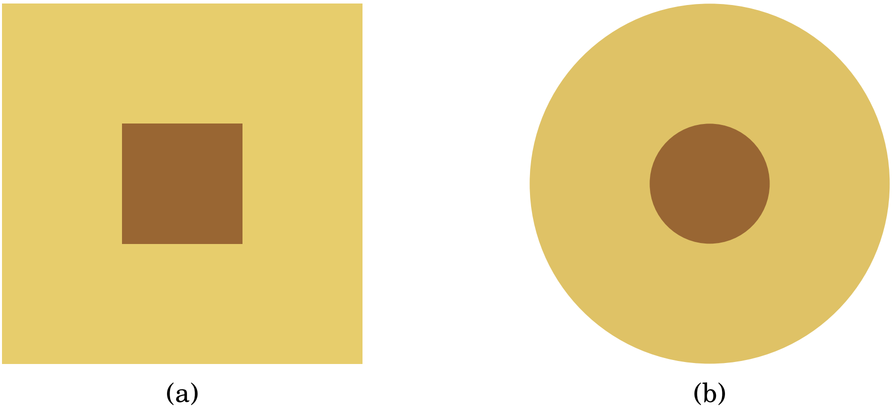

(a) (b)

Figure 3.3: The block structure for different p-norms. Beige points are 0-components and brown points are 1-components. (a) Block structure for $p = \infty$. (b) Block structure for $p = 2$.

also know that all 1-components are in $\odot_{\rho/2}\left(\vec{0}\right)$ (the brown region of in Figure 3.2). We also know that every process in the block is either 1-maintenance or 0-maintenance, so the product of every process is equal to its center-component.

One of the block's features that makes it especially useful is it exists for any p-norm, ρ, and D. Thus, Example 1 gives us a family of autopoietic systems that we can choose between, depending on our application. Note, however, that different p will result in different morphologies: $p = \infty$ will give a square and $p = 2$ will give a circle (Figure 3.3). The process neighbourhoods in these two organisations will similarly be squares and circles of radius ρ, respectively.

The block also makes clear that, in general, graphon organisations are very high dimensional objects. Counting up the dimensions of W_{Block}, we get $D+3$ from $K_0 = (\mathscr{N} \times \mathscr{M} \times \mathscr{M}) \cup \{0\}$ and D from Ω_{Block}, which gives a total of $D^2(D+3) = D^3 + 3D^2$. In $D = 2$ universes, W_{Block} is a $2^3 + 3(2^2) = 20$ dimensional object (a function over a 4-dimensional space whose values are 5-dimensional objects). Thus, visualising graphons is not generally feasible. But at least for the block, we can begin to get a grasp by reducing the dimensionality. If we ignore the 0 element and only label edges with $\mathscr{N}$, the values of this modified W are 2 dimensional, giving 8 dimensions in total. Taking this further, if we restrict ourselves to a boolean indicating whether any dependency relation exists at all, the values of

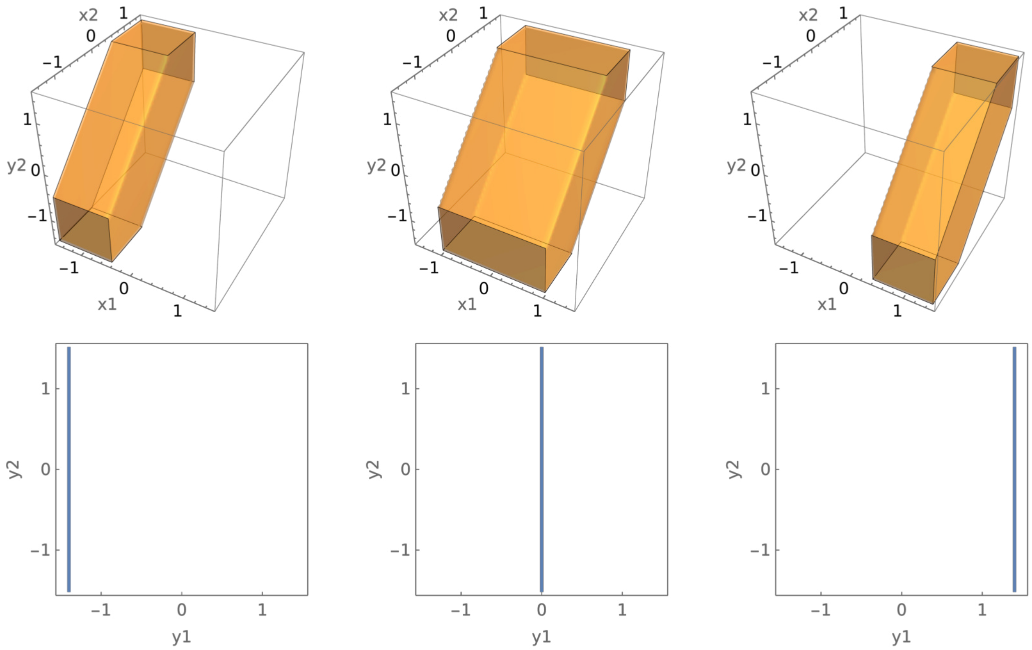


Figure 3.4: Simplified representation of the block graphon. $\langle x_1, x_2\rangle$ are the coordinates of source processes in Ω; $\langle y_1, y_2\rangle$ are the coordinates of target processes. On the bottom are line plots representing samples of target processes $\langle y_1, y_2\rangle$. Above are plots of the set points (x_1, x_2, y_2) such that $W(\langle x_1, x_2\rangle, \langle y_1, y_2\rangle) \neq 0$ for any $\langle y_1, y_2\rangle$ in the sample of the corresponding plot.

W become 1-dimensional objects, giving 5 dimensions in total, which we can visualise in a piecewise manner (Figure 3.4).

Example 2: $\rho \times 2\rho$ Blinker. Let

$$\begin{aligned}
\mathcal{B}_0 &:= \left[-\frac{3\rho}{2}, \frac{3\rho}{2}\right] \times \left[-2\rho, 2\rho\right] \times \{0\} \\
\mathcal{B}_1 &:= \left[-2\rho, 2\rho\right] \times \left[-\frac{3\rho}{2}, \frac{3\rho}{2}\right] \times \{1\} \\
\mathcal{B}_0^1 &:= \left[-\frac{\rho}{2}, \frac{\rho}{2}\right] \times \left[-\rho, \rho\right] \times \{0\} \\
\mathcal{B}_1^1 &:= \left[-\rho, \rho\right] \times \left[-\frac{\rho}{2}, \frac{\rho}{2}\right] \times \{1\}.
\end{aligned}$$

Let $\Omega_{Blinker} := \mathcal{B}_0 \cup \mathcal{B}_1$. The *blinker organisation* in $p = \infty$ universes is then defined

$$W_{Blinker}(x,y) := \begin{cases} \left(x - y, \mathbb{1}_{\mathcal{B}_i^1}(x), \mathbb{1}_{\mathcal{B}_{1-i}^1}(\langle 0,0,1\rangle - x)\right) & T(x,y) \\ 0 \end{cases} \tag{3.7}$$

where T is true when,

$$x \in \left[-\frac{3\rho}{2}, \frac{3\rho}{2}\right]^2 \times \{i\} \wedge y \in \mathcal{B}_{1-i} \wedge \|x - y\|_\infty \leq \rho. \tag{3.8}$$

$W_{Blinker}$ is reversible and not equivalent to its closure.

Example 2 appears much more complicated, but the logic of the definition is effectively the same as the block in Example 1. The main complication introduced here is the addition of a structure. This is why the sets $\mathcal{B}_i$ have the $\{i\}$ part – we need to separate processes by indexing by the structure they belong to. Thus $\mathcal{B}_0$ is the set of processes in the vertical blinker, and $\mathcal{B}_1$ is the set of processes in the horizontal blinker (Figure 3.5). Note also that the sets are aligned according to the transition the blinker undergoes. That is, $\mathcal{B}_0$ is centered on $\mathcal{B}_1$. The sets $\mathcal{B}_i^1$ are the 1-components of the corresponding blinker configurations $\mathcal{B}_i$ (Figure 3.5).

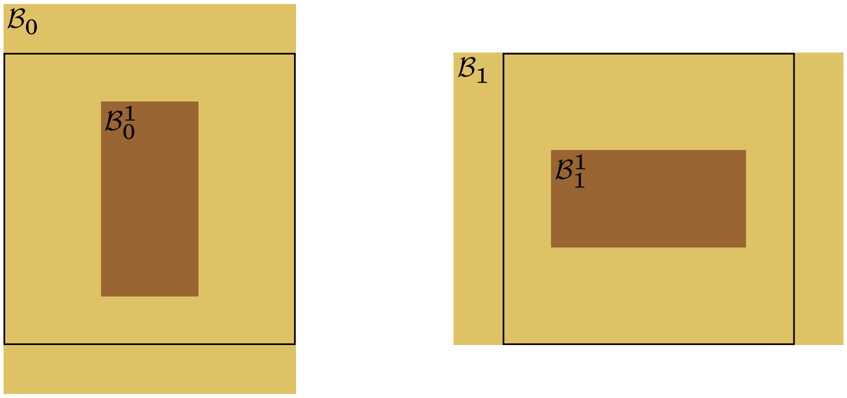


Figure 3.5: Ω representation of a $\rho \times (2\rho)$ blinker. $\mathcal{B}_0$ is the set of points in the vertical blinker configuration; $\mathcal{B}_1$ is the set of points in the horizontal configuration. $\mathcal{B}_i^1$ is the set of 1-components in the set $\mathcal{B}_i$. The black boxes outline the set of points that produce components in the organisation.

The edge decorations also change here to account for the fact that the product of a process in one blinker structure is always in the other structure. And since the sets $\mathcal{B}_i$ are aligned according to the structural transition, we get the product of a process at $x \in \mathcal{B}_i$ by finding the state of the component at the same point in the opposite blinker structure (this is why we need $\langle 0, 0, 1 \rangle - x$ to match the structure indices).

The condition T is also more simple than it may seem. The $||\cdot||_\infty$ part is the just same as in the block. The $y \in \mathcal{B}_{1-i}$ simply ensures that the enabled process is in the structure opposite of x. Finally, The $x \in [-\frac{3\rho}{2}, \frac{3\rho}{2}]^2 \times \{i\}$ ensures that only processes in the region where $\mathcal{B}_0$ and $\mathcal{B}_1$ overlap enable anything (Figure 3.5). This means that RealLife also contains processes that do not enable anything else, and processes that have no enabling relations pointing to their center-component. Thus, the degeneracy discussed in Chapter 2 where an organisation fails to specify a well-defined cognitive domain persists in RealLife, and the blinker exhibits this property. For instance, consider any $x \in \mathcal{B}_0 \backslash \left([-\frac{3\rho}{2}, \frac{3\rho}{2}]^2 \times \{0\} \right)$; such a process will always have $W(x, \cdot) \notin K$ and no process y where $W(y, x) = (\vec{0}, \cdot, \cdot)$.

Note that the morphology of this blinker is not a limit of the LtL blinkers analysed in Chapter 2. The LtL blinkers were characterised by a $\rho^* \times (\rho^* + 2)$ region of 1-components

($\rho^* \in \mathbb{N}$ denoting an LtL neighbourhood radius). But the limit of this is a block:

$$\lim_{\rho^* \to \infty} \left(\rho^* \frac{\rho}{\rho^*} \right) \times \left((\rho^* + 2) \frac{\rho}{\rho^*} \right) = \rho \times \rho$$

(the ρ/ρ^* factor is to treat ρ^* as a discretisation of ρ). Thus, the class of oscillators that converge on Example 2 are $\rho^* \times 2\rho^*$ blinkers.

This scaling problem also brings out another significant difference between RealLife and LtL: the neighbourhood radius does not qualitatively change anything in RealLife. In other words, the morphology of an emergent individual does not change with ρ as it does in LtL. This is why both Example 1 and Example 2 were parametrised by ρ.

As we will see, this organisation is quite degenerate. However, its organisation is a relatively simple example that has multiple structures and exhibits the relevant degeneracies that were observed in LtL.

Example 3: Translation-Invariant Bug. Let $\Omega_{Bug} := [0, 2\pi) \times (0, 1]$. Let $C^1 \subset \mathscr{S}$ denote the set of 1-components for a bug with stomach $S = (\mathscr{S}, \mathbf{s})$ with $p = 2$, and let $\vec{d}$ denote its translation vector. Let γ_0, γ_1, and $\gamma_2 = \gamma_1 + \rho$ be functions on $\mathbb{R}$ that represent the inner ∂C^1, outer ∂C^1, and $\partial \mathscr{S}$, respectively, in polar coordinates (Figure 3.6). Let $o(x) := \langle x_\theta,\ \gamma_2(x_\theta) x_r \rangle$ transform points in Ω to $\mathscr{S}$ in polar coordinates, where $x = \langle x_\theta, x_r \rangle \in \Omega$. Finally, let h be a function that transforms polar to Cartesian coordinates and let $\bar{h}(x, v) := h^{-1}(h(o(x)) - v)$ be a function that shifts a point x in polar coordinates by some Cartesian vector v. Then the organisation of a translation-invariant bug is defined

$$W_{Bug}(x,y) := \begin{cases} k(x,y) & T(x,y) \\ 0 & \end{cases} \tag{3.9}$$

where T is true when

$$\|h(o(x)) - h(o(y)) - \vec{d}\|_2 \leq \rho \ \wedge \ \bar{h}_r(x, \vec{d}) \leq \gamma_2(x_\theta) \tag{3.10}$$

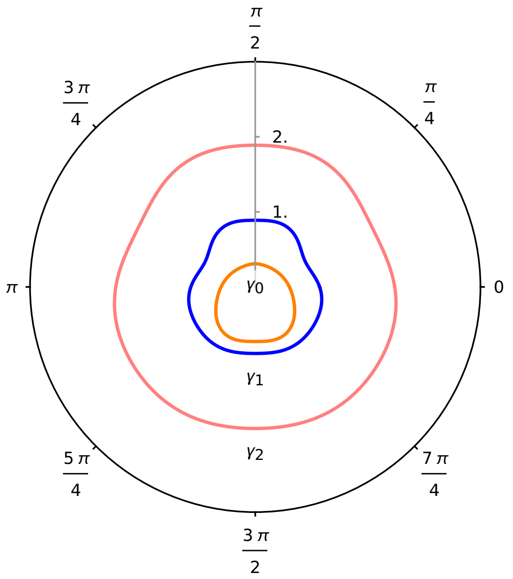


Figure 3.6: Representation of a bug structure by polar functions. In orange is the edge γ_0 of the bugs stomach (the inner region of 0-components). In blue is the outer edge γ_1 of the bugs 1-components. In pink is the outer edge γ_2 of the the bugs boundary.

and

$$\begin{aligned} k(x,y) = \Big(& h(o(x)) - h(o(y)) - \vec{d}, \\ & \mathbb{1}_{[\gamma_0(x_\theta),\gamma_1(x_\theta)]}\left(o_r(x)\right), \\ & \mathbb{1}_{[\gamma_0(\bar{h}_\theta(x,\vec{d})),\gamma_1(\bar{h}_\theta(x,\vec{d}))]}\left(\bar{h}_r(x,\vec{d})\right)\Big). \end{aligned} \tag{3.11}$$

W_{Bug} is reversible and not equivalent to its closure.

Example 3 is the most complicated organisation we will deal with, but most of its complexity simply arises from transitions between polar and Cartesian coordinates. The reason I use polar coordinates is to simplify the representation of smooth morphologies. For instance, we can get expressions for γ_0 and γ_1 as follows. First, we scale an LtL to some large ρ^* and find the boundaries of its region of 1-components (left side of Figure 3.7). Then we convert these boundaries to polar coordinates and linearly interpolate to get periodic functions $f_i : [0, 2\pi] \to \mathbb{R}$ (center of Figure 3.7). Finally, we define γ_i to be a Fourier

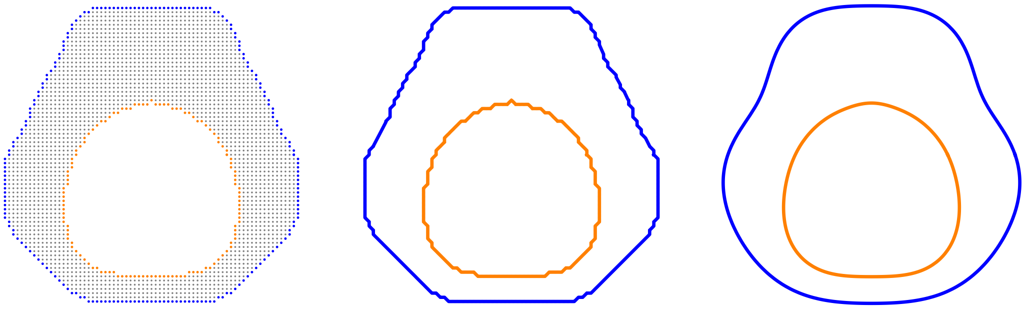


Figure 3.7: The derivation of a morphology from an LtL bug. On the left is a $\rho = 40$ LtL bug (the same as in Figure 2.3). 1-components are in grey, except for the two boundaries of the region, indicated in blue and orange. In the center is a linear interpolation of the blue and orange points in polar coordinates. On the right is a Fourier series approximation of the linearly interpolated functions (specifically, they 7 term Fourier series.)

series approximating the function f_i (right side of Figure 3.7). If the number of terms in the Fourier series (its accuracy) is relatively low, the morphology of the resultant bug will appear smooth – the series in Figure 3.7 and Figure 3.6 have 7 terms. Importantly, though, this procedure does not give the limiting morphology of the LtL bug, but just *a* morphology that is smooth and simple to describe mathematically. In fact, the the LtL bug used in Figure 3.7 is scaled from $p = \infty$ universes. If we wanted to approximate the limiting morphology properly, it would be best to scale an LtL bug from a $p = 2$ universe much further, and then use Fourier series with many more terms. (Note that Example 3 does not specify any specific morphology. That is, it describes the organisation of *any* translation-invariant bug whose structure can be represented with the polar γ functions.)

The reason why γ_2 is not defined in a similar manner is that the scaling method just described would, in general, fail to specify a structure with a proper boundary (the gap between γ_1 and γ_2 might not correspond to the neighbourhoods of the processes on γ_1). To avoid this problem, we assume $p = 2$ so that $\gamma_2 = \gamma_1 + \rho$ will result in a proper boundary. Thus, for bugs in RealLife with $p = \infty$, a different representation of their morphology will be needed.

These function now allow us to represent any translation-invariant bug with the same Ω, since we can simply scale every point in the structure by γ_2 (o is the inverse of this).

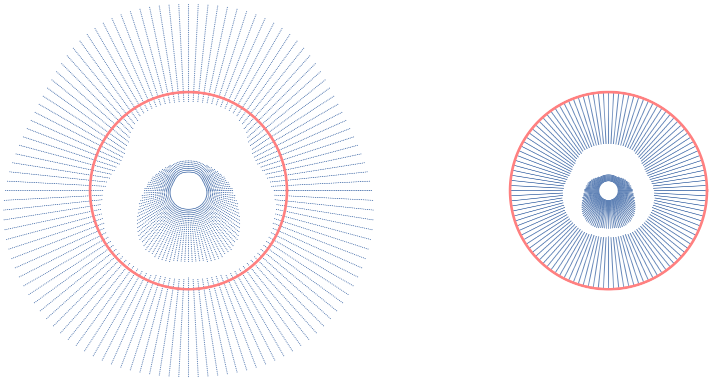


Figure 3.8: Scaling of a bug structure into Ω. On the left is a selection of 1-components in polar coordinates from the structure of the bug in Example 3. In red is the unit circle. On the right is the same set of points scaling by $\langle \theta, r \rangle \mapsto \langle \theta, r/\gamma_2(\theta) \rangle$. The red circle is still the unit circle, at the same scale as on the left.

This results in the transformation in Figure 3.8, where points in the structure are mapped to fit into Ω.

This Ω has a notable feature in that not every point in the structure is mapped. Namely the component at $\vec{0}$ is not contained in Ω. This is because such a point cannot be mapped uniquely into polar coordinates: $h : \langle 0, 0 \rangle \mapsto \langle \theta, 0 \rangle$ for all θ. However, this will not present a problem for us since we are assuming that any structures that are *almost everywhere* equal are good enough – we can have "holes" in Ω so long as the structures derived from the organisation are almost correct (in the technical sense).

With this, we can now understand the rest of the definition. The two terms in T correspond to whether (i) the process y, translated by $\vec{d}$, is in the neighbourhood of x; and (ii) whether x produces any component in structure. The edge decoration k is very similar to the blinker, except instead of accounting for processes in two different structures, we account for processes in different translations of a single structure. Thus, the first term of k gives the vector from the process y, translated by $\vec{d}$, to x (this needs to be expressed in Cartesian coordinates, since those are the coordinates of K). The next term gives the center-component of x by scaling it into the structure and determining whether it is in the region of 1-components between γ_0 and γ_1. The last term gives the product of x by shifting it by $-\vec{d}$ and determining whether this point is in the region of 1-components.

3.3. CONVERGENCE

In this section, I define in what sense process dependency graphons are the appropriate limiting objects of sequences of process dependency graphs. The proper notion of convergence we need is *homomorphism density* (Lovász, 2012, Chap. 17). The homomorphism density of a graph F into another graph G is the probability that a mapping of the vertices of F on to the vertices of G preserves the edges. Homomorphism densities can also be defined for graphons. This leads us to a major result of Graph Limit Theory: the limit of a sequence of graphs is the graphon such that the limit of the homomorphism densities is the homomorphism density of the graphon, for all F. More formally, a graphon W is the proper limit of a sequence (G_n) if for every simple graph F,

$$\lim_{n\to\infty} t(F, G_n) = t(F, W),$$

where t is the homomorphism density.

Unfortunately, defining t for process dependency graphs and graphons is not straightforward: there is no natural way to define the homomorphism density of one process dependency graph into another. However, what we can do is this. First, we scale enabling relations in LtL into RealLife, which we can do by simply multiplying the enabling relation by a scaling factor: $(r, \cdot, \cdot) \mapsto \left(r\frac{\rho}{\rho^*}, \cdot, \cdot\right)$. Thus, we treat LtL neighbourhoods as discrete approximations of a limiting RealLife neighbourhood. Then we can define the homomorphism density of a $\mathcal{Q}$-decorated graph into a process dependency graph, where $\mathcal{Q}$ is the space of real-valued functions on K. Finally, let $E^d_G(ij)$ give the edge decoration that the graph G assigns to the edge from i to j. Then the homomorphism density of a $\mathcal{Q}$-decorated graph F

into a scaled process dependency graph G_{ρ^*} is defined

$$t(F, G_{\rho *}) := \frac{1}{V(G_{\rho^*})^{V(F)}} \sum_{f:V(F)\to V(G_{\rho^*})} \prod_{ij\in E(F)} E^d_F(ij) \left(E^d_{G_{\rho^*}}(f(i)f(j)) \right). \tag{3.12}$$

The product here effectively assigns to each map f a value describing how well f preserves the edges. Thus, if any value in the product is 0 (i.e., an edge was not preserved), the value assigned to the map is 0. These values are then summed together and normalised by the number of maps. Note that t depends on a choice of decorations in $\mathcal{Q}$, and that this may result in densities not in $[0,1]$. Thus, the interpretation as the probability of an edge-preserving map does not in general hold here. Moreover, convergence must now exist over all F *and* $\mathcal{Q}$. Of course, it is impractical to simultaneously consider convergence under all possible functions in $\mathcal{Q}$ (an infinite dimensional space), but we can restrict ourselves to a subset $\mathcal{R} \subset \mathcal{Q}$ more amenable to calculation. Still, choosing this subset is a non-trivial task, but some natural choices will appear as we consider graphons.

To define homomorphism densities for graphons, we need to use the probability measures from ω, which amounts to a choice of $g : K_0 \to \mathbb{P}(K_0)$. The homomorphism density of a $\mathcal{Q}$-decorated graph F into a process dependency graphon ω is then defined

$$t(F, \omega) := \lambda \left[\Omega^{V(F)}\right]^{-1} \int_{\Omega^{V(F)}} \prod_{ij\in E(F)} \left[\int_{K_0} E^d_F(ij)(k)\,\omega(x_i, x_j)(dk) \right] \lambda[dx]. \tag{3.13}$$

The domain of the outer integral Ω^V is the space of maps from $V(F)$ into Ω, and x_i and x_j are the images of i and j under such a map. The inner integral is the value assigned to how the edge ij is mapped into ω. Note also that λ is not same as the D-dimensional measure used in Equation 3.1, but is instead a measure of the same dimension as $\Omega^{V(F)}$. — We can now state precisely the notion of convergence being used here: an infinite sequence of process dependency graphs (G_{ρ^*}) converges to a a graphon ω if, for every $\mathcal{Q}$-decorated graph F,

$$\lim_{\rho^*\to\infty} t(F, G_{\rho^*}) = t(F, \omega). \tag{3.14}$$

Choosing an appropriate g and $\mathcal{R}$ is in general a very difficult task. However, we can avoid this problem by using an alternative definition of t. Let $W_f(x,y) := \int_{K_0} f(k)\,\omega(x,y)(dk)$. If W_f satisfies certain properties for all $f \in \mathcal{Q}$ and every fixed $(x,y) \in \Omega^2$ (see Section A.3 in the Appendix for details), then Equation 3.13 can be rewritten as

$$t(F,\omega) = \lambda\left[\Omega^{V(F)}\right]^{-1} \int_{\Omega^{V(F)}} \prod_{ij\in E(F)} \left[W_{E_F^d(ij)}(x_i,x_j)\right] \lambda[dx]. \tag{3.15}$$

This allows us to compute t so long as we know the value of W_f for every $f \in \mathcal{R}$ (assuming F is decorated only with elements of $\mathcal{R}$). If we choose $\mathcal{R}$ to be a set of functions uniquely associated to each $k \in K_0$, we can design W_f as a function of two edge decorations in K_0 – one from $g^{-1} \circ E_F^d$ and one from W – that assigns a value to how well the enabling relations "align."

The simplest choice of W_f assigns 0 to all pairs of enabling relations that are not exactly equal. One way to do this is to assume $\mathcal{R}$ is the set of indicator functions of each element in K_0. Then we get

$$\begin{aligned} W_f(x,y) &= \int_{K_0} f(k)\,\delta(dk - W(x,y)) \\ &= f(W(x,y)) \\ &= \mathbb{1}_{\{k'\}}(W(x,y)), \end{aligned} \tag{3.16}$$

where $\delta(\cdot)$ is the δ-distribution over K_0, and $k' \in K_0$ is the enabling relation associated with $f \in \mathcal{R}$. A δ-distribution is defined to have the property $\int f(x)\,\delta(dx) = f(0)$ for all f. Thus, it is useful to use as a kind of filter to pick out the value of a function at specific points.

Unfortunately, this W_f is not very useful, as $t(\cdot,\omega)$ will usually be 0 for any ω. However, we can modify it to give non-zero values when two enabling relations $(r,c_0,c_1), (r',c_0',c_1') \in K_0$ satisfy $c_0 = c_0'$, $c_1 = c_1'$, and r "close enough" to r'. The only modification we need to make is to change $f \in \mathcal{R}$ to be an indicator function over some set. Formally, let

$\bigodot_a^K := \{x \in K_0 : \|x-k\|_p \le a\}$ give the intersection of K_0 and the ball of radius $a < 1$ surrounding a point $k \in K_0$. Now we get:

$$\begin{aligned} W_f(x,y) &= \int_{K_0} f(k)\,\delta(dk - W(x,y)) \\ &= f(W(x,y)) \\ &= \mathbb{1}_{\bigodot_a^K(k')}(W(x,y)) \end{aligned} \tag{3.17}$$

where k' is again the point in K_0 associated with $f \in \mathcal{R}$. This definition will give 1 when the decorations on F and ω are within a of each other and share the same c_0 and c_1, and will give 0 otherwise.

To make this more concrete, we can compute homomorphism densities for sequences of LtL blocks and the block in Example 1. But note that most $\mathcal{R}$-decorated graphs F will give $t(F,\cdot) = 0$. For instance, if F decorates any edge with (r, c_0, c_1) where $c_0 \neq c_1$, no pair of vertices in the block will preserve that edge, since all processes in the block are either 0-maintenance or 1-maintenance. Thus, to observe non-trivial converge, we have to be more careful about what graphs we choose. A simple choice is a graph F_1 with a single vertex and $E^d_{F_1} = \mathbb{1}_{\{(\vec{0},1,1)\}}$. In this case, we can solve Equations 3.12 and 3.15 analytically, by noticing that every map of F_1 into G or ω picks out a single process. Thus, we can reduce Equation 3.12 to

$$t(F_1, G_{\rho^*}) = \frac{1}{V(G_{\rho^*})} \sum_{i \in V(G_{\rho^*})} E^d_{F_1}\left(E^d_{G^*_\rho}(ii)\right),$$

which will give the proportion of 1-maintenance processes in $V(G_{\rho^*})$. Assuming $p = \infty$, we can compute this by

$$t(F_1, G_{\rho^*}) = \frac{(\rho^* + 1)^2}{(3\rho^* + 1)^2}$$

which gives rise to the limit

$$\lim_{\rho^* \to \infty} t(F_1, G_{\rho^*}) = \lim_{\rho^* \to \infty} \frac{(\rho^* + 1)^2}{(3\rho^* + 1)^2} = \frac{1}{9}. \tag{3.18}$$

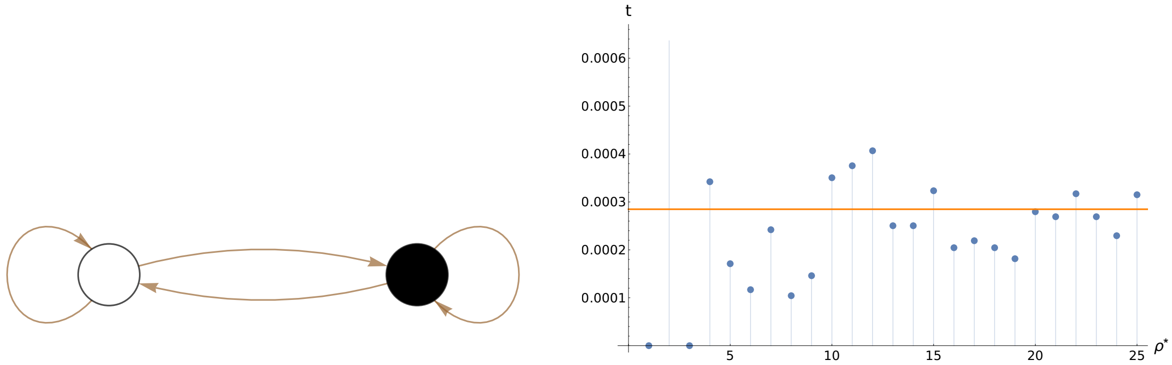
Similarly for the graphon ω_{Block}, Equation 3.15 becomes

$$\begin{aligned} t(F_1, \omega) &= \lambda[\Omega]^{-1} \int_\Omega W_{E^d_{F_1}}(x,x) \lambda[dx] \\ &= \frac{1}{9\rho^2} \int_\Omega \mathbb{1}_{\odot^K_a((\vec{0},1,1))}(W(x,x)) \lambda[dx] \\ &= \frac{1}{9\rho^2} \rho^2 = \frac{1}{9}, \end{aligned}$$

which is obviously the same result as Equation 3.18. By the same logic, one can also show that if F were decorated with $(\vec{0}, 0, 0)$ instead, we would get the limit $\frac{8}{9}$.

More generally, we can use single vertex graphs like these to show that a graphon has the limiting proportion of processes of each type. Note, however, that none of this constitutes *proof* that the block graphon defined in Example 1 is the limiting object of LtL blocks, but it does provide compelling evidence that this should be the case (again, a proof would require us to show convergence for *every* $\mathcal{Q}$-decorated graph F).

For less trivial graphs, we will usually have to compute homomorphism densities numerically. Beaujean et al. (2021) describe an algorithm to compute homomorphism densities for simple graphs by random sampling, which can be extended quite easily to accommodate decorated digraphs. Similarly, we can compute Equation 3.15 by numerical integration. Let F_2 be a graph with two vertices, shown in Figure 3.9a. The two self-connections are labelled with indicator functions centered at $(\vec{0}, 1, 1)$ and $(\vec{0}, 0, 0)$ for the white and black nodes, respectively. The cross-connections are labelled with indicator functions centered at $(\langle 0, \frac{\rho}{2} \rangle, 1, 1)$ for the black-to-white edge and $(\langle 0, -\frac{\rho}{2} \rangle, 0, 0)$ for the white-to-black edge. That is , F_2 is a graph containing a 1-maintenance and 0-maintenance process with a reciprocated pair of edges between them. Figure 3.9b shows the convergence of $t(F, G_{\rho^*})$



(a) (b)

Figure 3.9: Homomorphism densities of a two-node graph into block organisations. (a) A $\mathcal{R}$-decorated graph F with two vertices and four edges. (b) A sequence of homomorphism densities of the graph in (a) into LtL blocks of increasing ρ^*. Blue dots are the values of $t(F, G_{\rho^*})$ for the $(\rho^* + 1) \times (\rho^* + 1)$ block organisation G_{ρ^*}. The orange line is the value of $t(F, \omega_{Block})$ for $\rho = 1$.

for a sequence of $(\rho^* + 1) \times (\rho^* + 1)$ LtL blocks in blue, and $t(F, \omega_{Block})$ in orange, with $a = \frac{1}{10}$. Clearly, the convergence is non-monotonic, but it still approaches the density computed for the graphon.

There is an important point regarding homomorphism densities that restricts the space of appropriate graphon limits. Namely, for a sequence of graphs (G_i) with interaction graphs $\mathcal{I}_i = (\mathcal{S}_i, \mathcal{T}_i, \mathfrak{s}_i, \mathfrak{t}_i, \tau_i)$, the total area of all structures in $\mathcal{S}_i$ should converge to the area of Ω. The block in Example 1 and the blinker in Example 2 both exhibit this property, but the bug in Example 3 does not. By forcing processes to fit into the unit circle (Figure 3.8), the area of any subset of the structure is not preserved. The simplest way to avoid this is to define Ω implicitly as the area bounded by γ_2 (the outer most function in Figure 3.6). This would make the homomorphism density more difficult to calculate by integrating over an irregular $\Omega^{V(F)}$. The area of Ω would also have to be computed differently – for example, by a polar integral when using a polar Fourier series representation:

$$\lambda[\Omega] = \frac{1}{2} \int_0^{2\pi} \gamma_2 (\theta)^2 \, d\theta.$$

In sum, then, we have evidence that, just as structures can be scaled from LtL to Real-

Life, so can organisations, where the limit of process dependency graphs of increasing ρ^* is a process dependency graphon. This connection permits us to approximate graphons by graphs and vice versa.

3.4. The Cognitive Domain

Due to our assumption of organisations with finitely many structural realisations, we can represent the cognitive domain in RealLife by an in interaction graph, just as in LtL. Formally, we again define an interaction graph $\mathcal{I} = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$, where $\mathcal{S}$ is a set of structures $S = (\mathscr{S}, \mathbf{s})$, $\mathcal{T}$ is a set of edges $(S_i, S_j, \vec{v})$ describing transitions between structures, $\mathfrak{s}$ and $\mathfrak{t}$ map edges to the source and target vertex, respectively, and $\tau : \mathcal{T} \to \mathcal{C}$ labels edges with perturbation classes. I preserve the multigraph definition from LtL for sake of consistency, even though no graphon will have a multigraph cognitive domain.

The general procedure for deriving the interaction graph of a graphon ω is composed of the same two basic steps as in LtL: *extracting* and *merging* fragments. The main difference here is that we cannot extract the fragments of every process before merging, since we now have infinite process sets. Similarly, we will have to merge fragments without knowledge of how every process is embedded. To overcome these difficulties, we reconstruct one structure at a time by alternating between extraction and merging steps, keeping track of how a finite set of processes are embedded in the fragments. Unfortunately, this requires a great deal of mathematical machinery to work in the general case, but major simplifications can be made when the organisation is reversible and the structures are sufficiently small.

We denote the *forward fragment* extracted from a process $x \in \Omega$ by $\vec{F}_x = (\overrightarrow{\mathscr{F}_x} \subset \mathbb{R}^D, \vec{\mathbf{F}}_x : \overrightarrow{\mathscr{F}_x} \to \mathscr{M})$ (clearly, this has the same form as a structure $(\mathscr{S}, \mathbf{s})$, and in fact the goal of this procedure is to modify $\vec{F}_x$ until it becomes a well-defined structure). To define $\vec{F}_x$, let $\overrightarrow{W}^z(x) := \{y : W_r(x,y) = (z, \cdot, \cdot)\}$ denote the set of processes that x produces the z reactant of; then:

$$\overrightarrow{\mathscr{F}}_x := \bigcup_{z \in \mathscr{N}} \phi_{\overrightarrow{W}^z(x)} - z \tag{3.19}$$

$$\overrightarrow{\mathbf{F}}_x := \bigcup_{z \in \mathscr{N}} \sigma^{-z} \circ \mathbf{p}_{\overrightarrow{W}^z(x)}. \tag{3.20}$$

These are the RealLife analogues of Equations 2.5 and 2.6 from the previous chapter: we are taking the translated union of processes in in the organisation. Note, however, that this is only well defined for reversible organisations, since $\overrightarrow{W}$ need not give a single process in the general case (x may produce more than one z reactant in a non-reversible organisation). Thus, we need a way to separate the processes enabled by x into a finite set of fragments. This is where a major limitation of the theory arises: *there is no clear way to separate fragments without additional knowledge of how processes are arranged in* Ω. This is not to say that it is necessarily impossible to separate fragments, but that it is at the very least quite difficult. To proceed with the theory, I will assume some knowledge of Ω, but only to extract fragments – the rest of the procedure should hold regardless of these assumptions. Of course, this still violates the point made above that the cognitive domain should be derived without assuming such knowledge of Ω. Thus, it is an open problem whether it is possible to forgo these assumptions in non-reversible organisations (again, no assumptions are necessary when the organisation is reversible).

One way to separate fragments is to assume knowledge of a partition $\{O_i\}$ of Ω into structures. Note that this partition need not tell us anything about how processes are arranged within each structure, just what structure the process belongs to. Applied to the examples in Section 3.2, the block and bug's partition would be $\{\Omega\}$ and the blinkers partition would be $\{\mathcal{B}_0, \mathcal{B}_1\}$ (of course, these are just examples: none of these organisations *require* the partition, since they are all reversible). Then we can extract a set of fragments

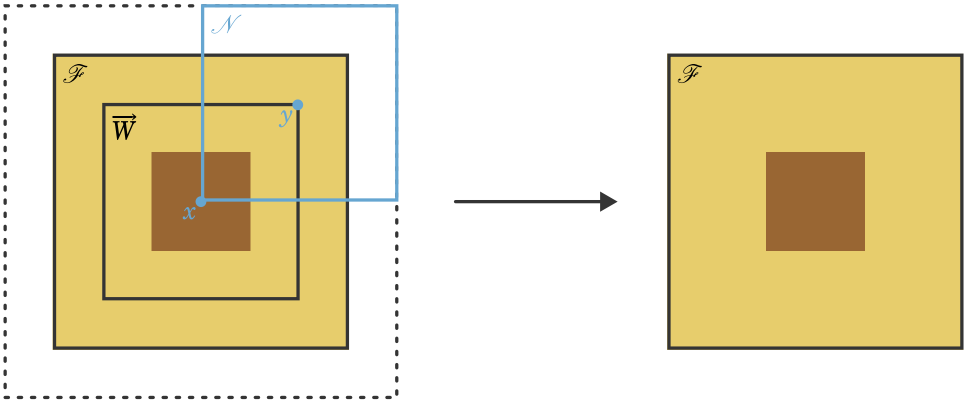


Figure 3.10: A fragment extracted from the block. x is the center-most process in the block. $\overrightarrow{W}$ is the set of processes x enables. y is a process enabled by y, whose neighbourhood (in blue) extends beyond the block. The dashed line is the union of all the neighbourhoods of each process in $\overrightarrow{W}$, which is the largest possible fragment that can be extracted for any organisation. $\mathscr{F}$ is the actual fragment extracted from x, i.e., the union of ϕ_y for all $y \in \overrightarrow{W}$.

from $x \in \Omega$ by indexing by the partition:

$$\overrightarrow{\mathscr{F}}_{x,i} := \bigcup_{z \in \mathscr{N}} \phi_{\overrightarrow{W}^z(x) \cap O_i} - z \tag{3.21}$$

$$\vec{\mathbf{F}}_{x,i} := \bigcup_{z \in \mathscr{N}} \sigma^{-z} \circ \mathbf{p}_{\overrightarrow{W}^z(x) \cap O_i}. \tag{3.22}$$

Figure 3.10 demonstrates a fragment extraction in the block for $x = \vec{0} \in \Omega$. Intuitively, we can break down this extraction into two steps: embedding the processes, and taking the union of their functions. The embedding here is visualised by the box labelled $\overrightarrow{W}$, as these are all the processes enabled by x. Clearly, the actual fragment is bigger than this embedding. For example, translating $\mathscr{N}$ to be centered on the process y, we see that, in general, the largest possible fragment extracted from a process is the p-norm ball of radius 2ρ: a $4\rho \times 4\rho$ box when $p = \infty$ (the dashed line in Figure 3.10). Thus, when the entire structure is smaller than this (as in the block), an appropriately chosen process will extract a fragment corresponding to the complete structure.

Of course, we cannot make this assumption in the general case. Thus, the next step of the procedure is to *expand* the fragment until it becomes a well-defined structure. The basic idea is this: given a fragment $\vec{F}$ and a well-chosen process $x \mapsto x' \in \overrightarrow{\mathscr{F}}$, we find

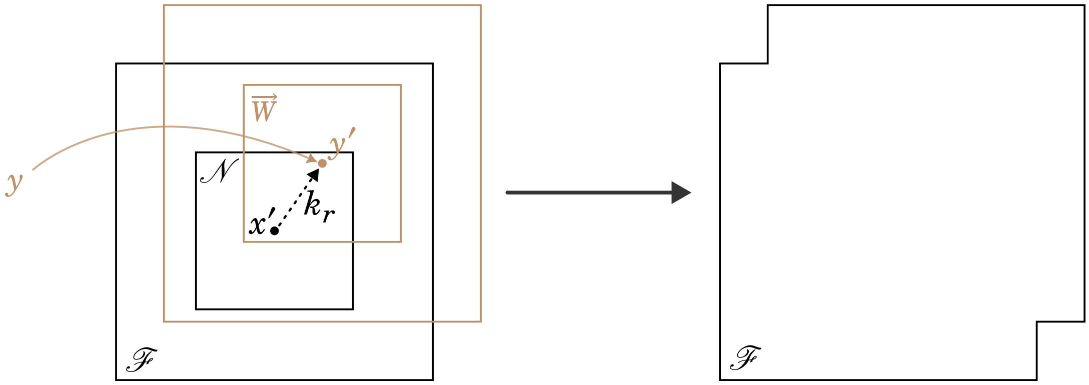


Figure 3.11: Fragment expansion procedure. Given an $x' \in \mathscr{F}$ enabled by y such that $W(y,x) = (k_r,\cdot,\cdot)$, the point the forward fragment of y (largest brown square) can be merged into $\mathscr{F}$ at the point y'. The square $\overrightarrow{W}$ is the set of processes that y directly enables. The larger brown square is the union of all of those processes. This expansion process results in the single larger fragment on the right.

the processes $\{y_i\}$ that enables x by some k and merge one of their fragments $\vec{F}_{y_i}$ into $\vec{F}$, displacing it by k_r (Figure 3.11). Clearly, if $k_r \neq \vec{0}$ and x' is not near the boundary of the structure, this will expand the fragment. What remains is to choose an appropriate x and values for k_r. We can achieve both as follows. For the fragments extracted from each process x, keep track of the pairs (x',x''), where $W(x,x'') = (x' = \vec{0},\cdot,\cdot)$; if the fragment is merged to another by a displacement v, modify the point $(x',x'') \mapsto (x'+v,x'')$. Then, to expand a fragment containing the process x_i'', select some number of $r \in \partial\mathscr{N}$ (for instance, one in each corner when $p = \infty$) and expand the fragment as above, merging onto $x_i' + r$. Thus, for each such set of expansions, we get a new set of points (x',x''), allowing us to continue expanding around each $x' \in \overrightarrow{\mathscr{F}}$.

The expansion procedure can be terminated when every merge results in the same fragment. Note that we only have to keep track of finitely many processes for this to work, since any pair of fragments to be merged will generally overlap over most of their respective domains. However, there is a degenerate case here that carries over from LtL: if an organisation contains processes that have no enabling relations pointing to their center component *and* that do not enable any process, then this procedure will fail. Just as in LtL, components can only be specified by edges, and the center-component of these de-

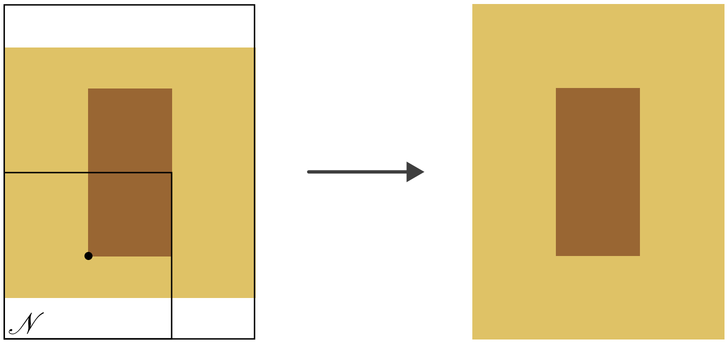


Figure 3.12: Recovery of missing components from degenerate blinker organisation. The black dot is a 1-component whose neighbourhood contains some of the missing components. By assuming that all components in the neighbourhood of such processes not already specified are 0, one recovers the full blinker structure on the right.

generate processes are never described by any edge. Moreover, we do not have the luxury here of being able to specify processes explicitly, due to the infinite process sets we require. Thus, the blinker organisation defined in Example 2 fails to derive a well-defined cognitive domain: the derived structures will be missing boundary components. Practically, this degeneracy can be overcome by "filling in" where the derivation fails to specify components, by adding in the 0-components in the neighbourhoods of the 1-components (Figure 3.12) – we can guarantee that degenerate processes are necessarily in the boundary of the structure. But this is theoretically unsatisfying, as the organisation no longer specifies that these components should exist, even though its persistence depends on them – there is no complete map from processes to their spatial embedding implicit in the organisation here.

Returning to the general procedure, there are still additional steps that need to be taken to determine whether the organisation has more than one structural realisation and, if it does, to derive the remaining structures. Given a set of completed fragments $\{S_i\}$ (now *structures*), the first step is to extract the forward fragments from each of the points kept during the expansion. Then, if none of these fragments can be merged into an existing structure, those fragment belong to a new structure. To determine whether a fragment $\vec{F}_x$ can be merged into the structure S_i, we need to find a process in S_i that x enables. This can be done as follows. First, find a vector v such that when $\vec{F}_x$ is displaced by v, it is contained within $\mathscr{S}_i$ and agrees with $\mathbf{s}_i$ over $\overrightarrow{\mathscr{F}_x} + v$. Formally, this amounts to finding a $v \in \mathscr{S}_i$ such that

$$\epsilon_v\left[\vec{F}_x, S_i\right] = \vec{\mathbf{F}}_x\Big|_{\mathscr{S}_i\cap(\vec{\mathscr{F}}_x+v)} - \mathbf{s}\Big|_{\mathscr{S}_i\cap(\vec{\mathscr{F}}_x+v)} = 0. \tag{3.23}$$

Then, if this v exists, one can find an enabling relation from x into S_i if, for any $y'' \in \mathscr{N} + v \cap \mathscr{S}_i$, $W(x,y') \neq 0$ where (y',y'') is one of the points tracked during the expansion of S_i. Recall that every expansion involved overlapping neighbourhoods (Figure 3.11). Therefore, every point in $\mathscr{S}_i$ must be in the neighbourhood of some tracked (y',y''). If $\vec{F}_x$ *cannot* be merged into any existing structure, it should be used to start the expansion process of a new structure.

Now we can start putting these structures in the form of the interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ of the process dependency graphon ω. Fortunately, at this point most of the procedure is just as in LtL. We get the structure set $\mathcal{S}$ by centering all of the derived structures around $\vec{0}$. We get the edge set $\mathcal{T}$ by considering all pairs of structures in $(S_i, S_j) \in \mathcal{S}^2$ and determining whether any process in S_i enables any process in S_j, using the merging method just described above. Specifically, we get the edge (S_i, S_j, v) by finding any pair (x_i', x_j') such that $W(x_i', x_j') = (-v, \cdot, \cdot)$, where $(x_i',\, x_i'' \in \mathscr{S}_i)$ and $(x_j',\, x_j'' \in \mathscr{S}_j)$ are tracked points for S_i and S_j, respectively (note that this requires the tracked points to also be shifted according to the centering of the structures).

The perturbation class function $\tau : \mathcal{T} \to \mathcal{C}$ is derived just the same as in LtL: assign any pair of edges the same label in $\mathcal{C}$ if the same transformation (rotation and/or reflection) can be applied to both structures in one edge to get the structures in the other (while accounting for the displacement vectors). Formally, if $t_1 = (S_{1i}, S_{1j}, v_1)$ and $t_2 = (S_{2i}, S_{2j}, v_2)$ are edges in $\mathcal{T}$, then $\tau(t_1) = \tau(t_2)$ when there exists some transformation $\mathbf{f}$ such that $\mathbf{f} \circ \mathbf{s}_{1i} = \mathbf{s}_{2i}$ and $\mathbf{f} \circ (\sigma^{v_1} \circ \mathbf{s}_{1j}) = \sigma^{v_2} \circ \mathbf{s}_{2j}$.

Interaction graphs in RealLife look very much the same as in LtL, except now the structures are objects in continuous space (Figure 3.13). Thus, all of the general interpretation and utility discussed in Section 2.4 holds just as well here.

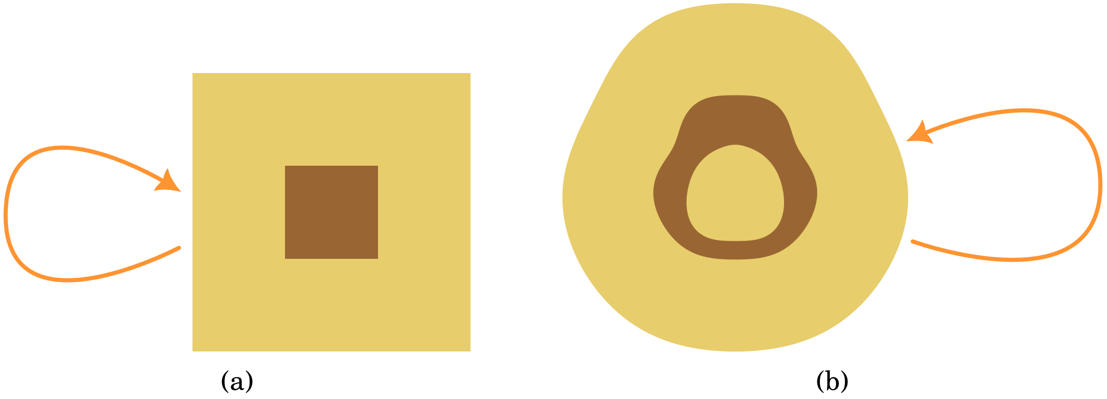

(a) (b)

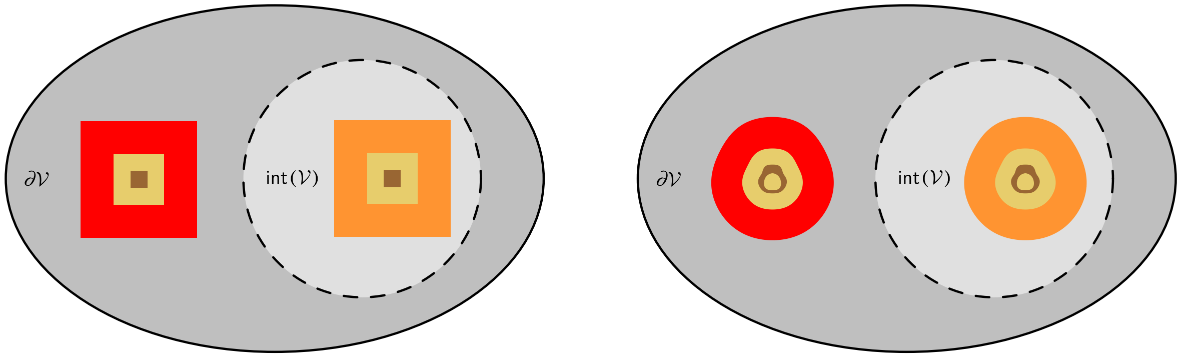
Figure 3.13: Interaction graphs for the block and translation-invariant bug. Structures are drawn to scale for the same ρ. (a) The self-connection here represents an edge $(S, S, \vec{0})$. (b) The self-connection here represents an edge $(S, S, \vec{d})$, where $\vec{d}$ is the translation vector of the bug.

3.5. The Viability Constraint

The procedure for deriving the intrinsic viability constraint from a process dependency graphon is nearly identical to the procedure for LtL described in Section 2.5. The only differences arise from the exchange of discrete structures for continuous structures, but this introduces nothing fundamentally different to the definitions (though some features of the resultant objects change). Thus, I describe generalisations of both the *state formulation* and the *density formulation*.

3.5.1. The State Formulation

The state formulation is again quite simple: given an interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$, the set of viable (1,2)-environment configurations $\mathcal{V}_\omega$ is defined as the set of functions over $\mathcal{E}^2(\mathscr{S}_i)$ that contain the structure in $S_i \in \mathcal{S}$. The boundary of the viability constraint $\partial\mathcal{V}_\omega$ is then defined as the subset of $\mathcal{V}_\omega$ where the updated function does not contain a structure. Then $\mathsf{int}(\mathcal{V}_\omega) := \mathcal{V}_\omega \backslash \partial\mathcal{V}_\omega$.



(a)

(b)

Figure 3.14: Intrinsic viability constraints of a block and bug organisation. The red (1,2)-environment represents the class of destructive configurations, while orange represents the class of environments that preserve the structure.

Example 4: The Block. The intrinsic viability constraint of the block (as defined in Example 1) is given by

$$\mathcal{V}_{Block} := \mathcal{E}^2_\Omega \left[\mathbb{1}_{\odot_{\rho/2}(\vec{0})} \right] \tag{3.24}$$

$$\partial\mathcal{V}_{Block} := \left\{ \mathbf{u} \in \mathcal{V}_{Block} : \xi[\mathbf{u}]\Big|_\Omega = \mathbb{1}_{\odot_{\rho/2}(\vec{0})} \right\}, \tag{3.25}$$

where (by a slight abuse of notation) $\mathcal{E}^n_{\mathscr{A}}[\mathbf{u}]$ denotes the space of functions over $\mathcal{E}^n(\mathscr{A})$ that are equal to $\mathbf{u}$ over $\mathscr{A}$. The viability constraint is depicted schematically in Figure 3.14a.

The main difference from LtL here is that these sets are now *infinite dimensional* functions spaces. Thus, searching this space for perturbations to induce particular transitions becomes extremely prohibitive. Hence we see that, even more than in LtL, the density formulation will be is very useful here.

3.5.2. The Density Formulation

The density formulation expresses the set of viable (1,2)-environments as density functions over structures and their 1-environments. Again, we will need some more mathematical machinery, analogous to the definitions we used in LtL.

First, the core of the formulation – the inverse relation between a universe state function $\mathbf{u}$ and its density function Ψ – depends on results from Fourier analysis (Folland, 1992, pp. 206–244; also see Section A.2.3 in the Appendix). Thus, we get

$$\mathbf{u} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\Psi]}{\mathcal{F}[\kappa]}\right], \tag{3.26}$$

where $\mathcal{F}$ is an appropriate multidimensional Fourier transform, and κ is the convolution kernel defined above (Equation 3.1). To get the density bounds at x for a source structure $\mathbf{a}$ and target structure $\mathbf{b}$, we define

$$\Gamma[\mathbf{a},\mathbf{b}](x) := \begin{cases} [b_0, b_1] & \mathbf{a}(x) = 0 \wedge \mathbf{b}(x) = 1 \\ [0,1] \backslash [b_0, b_1] & \mathbf{a}(x) = 0 \wedge \mathbf{b}(x) = 0 \\ [s_0, s_1] & \mathbf{a}(x) = 1 \wedge \mathbf{b}(x) = 1 \\ [0,1] \backslash [s_0, s_1] & \mathbf{a}(x) = 1 \wedge \mathbf{b}(x) = 0. \end{cases} \tag{3.27}$$

(Clearly, this depends on a choice of a RealLife rule). For any edge $(S_i, S_j, v) \in \mathcal{T}$, $\Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j]$ then gives the family of density functions over $\mathscr{S}_i \cap (\mathscr{S}_j + v)$ such that $\xi[\mathbf{s}_i](x) = \mathbf{s}_j(x)$ for all x in that set (the violet region in Figure 3.15). To get the family of density functions over the rest of $\mathscr{S}_j + v$ (orange regions in Figure 3.15), we apply Equation 3.26 to any valid density function satisfying Γ to get a set of components over all of $\mathscr{S}_j + v$. Then a family of density functions can be defined over all of $\mathscr{S}_j + v$, each such function implying a unique set of components relevant to the transition.

This allows us to define $\mathcal{V}$ as the set of all density functions Ψ such that Equation 3.26 gives the function $\mathbf{s}_i(x)$ for all $x \in \mathscr{S}_i$ for some edge $(S_i.S_j, v) \in \mathcal{T}$ of the interaction graph. $\mathtt{int}(\mathcal{V})$ then becomes the subset of these functions with $\Psi(x) \in \Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j]$ for all $x \in \mathscr{S}_j + v$, for some edge $(S_i.S_j, v)$. There are a few notable features of this definition worth mentioning here. First, $\mathcal{V}$ is still an infinite dimensional space of functions, except instead of needing to specify a value in $\mathscr{M}$ for every point in $\mathcal{E}^2(\mathscr{S}_i)$, we specify a value in $[0,1]$ for every point in $\mathscr{S}_j + v$.

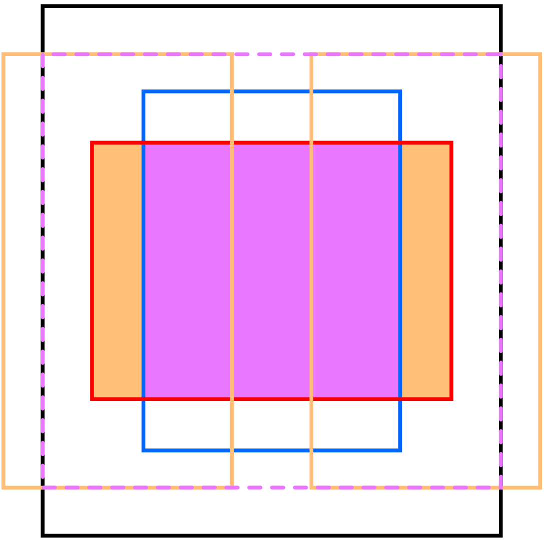


Figure 3.15: Schematic of a structural transition. The blue outline represents the initial configuration, and the red outline represents the configuration after updating. The black outline represents the 1-environment of the blue configuration. The violet region is the intersection of the red and blue configurations; the dotted violet line is the 1-environment of this region. The orange regions are the intersection of the red configuration with the 1-environment. The orange outlines are the 1-environments of the orange regions.

Second, a density function Ψ can be tangent to the upper or lower bounds of Γ while still maintaining viability. In order for Ψ to violate the viability of the organisation, Ψ must be outside a given Γ over some region with non-zero area (assuming that no other Γ is satisfied). Thus, we say that a perturbation destroys the organisation when it makes a *measurable difference* to the density function Ψ with respect to all Γ. Interestingly, if we assume that all universe state functions $\mathbf{u} \in \mathcal{U}$ are absolutely integrable (which we have been) and $\Psi = \kappa * \mathbf{u}$ is only ever tangent to the bounds of Γ over sets of zero area (which is true for most of $\mathcal{U}$), then ξ is continuous on $\mathcal{U}$ (Pivato, 2007, Theorem 1.1). In other words, slight variations in $\mathbf{u}$ or Ψ will induce slight variations in $\xi[\mathbf{u}]$.

Third, density functions are continuous but not (generally) differentiable (Figure 3.16). This last point is especially relevant to specifying when a density function has an inverse in $\mathcal{U}$ – i.e., not every $\Psi : \mathscr{A} \to \mathbb{R}$ has a universe state function as an inverse, so it would be useful to determine when Ψ has such an inverse purely in terms of operators on density functions. However, the standard definition of the partial derivative of a convolution –

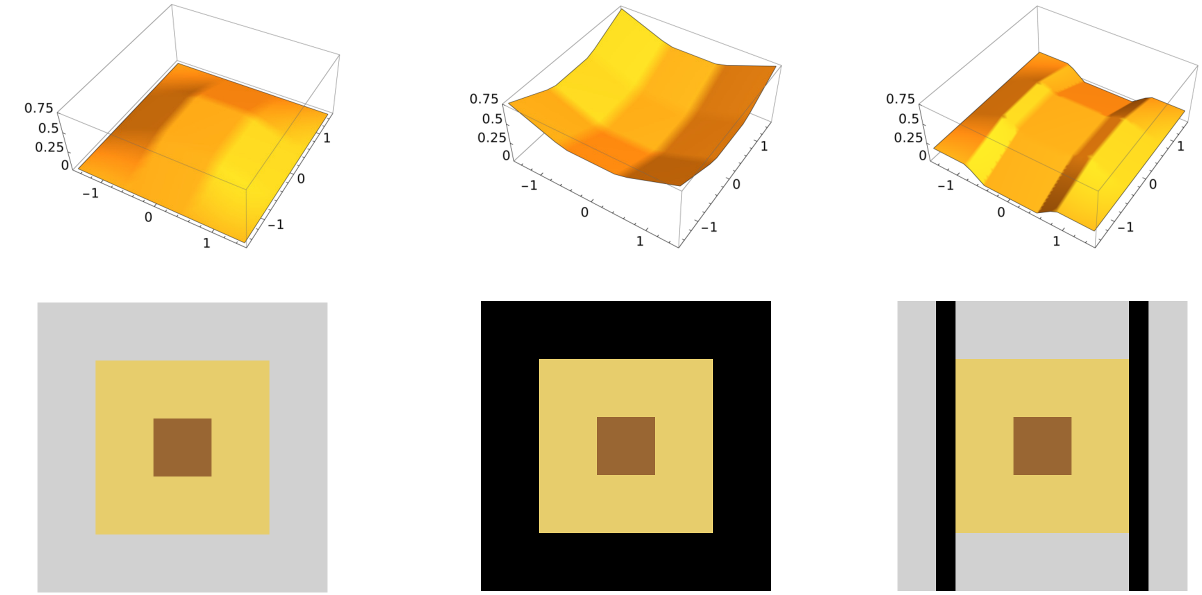


Figure 3.16: Continuous density functions over the block. The bottom row contains blocks and 1-environments, where beige and brown points indicate the 0- and 1-components belonging to the block, respectively, and grey and black points are the 0- and 1-components in the 1-environment. On top is the density function over the block structure for the $\mathcal{E}(\mathscr{S})$ configuration depicted below.

$\partial(f * g) = (\partial f) * g$ – is either 0 or undefined for indicator functions, but clearly the rate of change of a density function is not always equal to 0 (Figure 3.16). It is possible to get the correct derivative with an appropriate analytic expression of Ψ directly in terms of $x \in \mathbb{R}^D$. Alternatively, one can take advantage of a convenient property of Fourier transforms: $\mathcal{F}[\partial_{x_i} f](x) = -ix\mathcal{F}[f][x] \implies \partial_{x_i}\Psi(x) = -i\mathcal{F}^{-1}[x\mathcal{F}[\Psi](x)]$. Thus, we can compute the derivative of Ψ in terms of its Fourier transform.

The topology mentioned above works to guarantee that continuous deformations to density functions will correspond to continuous deformations of the structure after updating (under a very specific sense of continuous). Thus, it may seem that we should be able to use the metric defining that topology to find perturbations on a structure S that do or do not preserve the organisation. But it turns out that that metric is not very useful for this, as it tells us nothing about *where* a perturbation should be changed. Thus, we will need a more organisation-relative metric to achieve this spatial resolution, extending the metric defined in Section 2.5. Define:

$$d_t[S_i, \Psi](x) := \inf\left\{|\Psi(x) - y| : y \in \Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j](x)\right\}, \tag{3.28}$$

where $t = (S_i, S_j, v) \in \mathcal{T}$ and the inverse of Ψ gives $\mathbf{s}_i$. Just as in LtL, this metric gives the distance Ψ is from being tangent to a bound of Γ at x. This metric can be extended to give the distance of Ψ from any bound of any Γ:

$$d_{\mathcal{V}}[S_i, \Psi](x) := \min\left\{d_t[S_i, \Psi] : t = (S_i, \cdot, \cdot) \in \mathcal{T}\right\}. \tag{3.29}$$

Using either of these metrics, one can apply continuous deformations to density functions at every $x \in \mathcal{E}(\mathscr{S}_i)$ until the desired transition, determined by Γ, is achieve at those points. Note that if Ψ is changed in $\mathscr{S}_i$, Γ will need to be recomputed over the 1-environment. This problem can be avoided, though, so long as one focuses on a single transition and splits the task into two repeated steps: (i) find a valid Ψ over $\mathscr{S}_i \cap (\mathscr{S}_j + v)$ and compute Γ over the rest of $\mathscr{S}_j$; and (ii) continuously deform a valid extension Ψ' until $d_t[S_i, \Psi'](x) = 0$ for all $x \in \mathscr{S}_j + v$ and $\Psi' = \Psi$ over the initial region. If no Ψ' satisfies Γ, get a new Ψ by changing the components over $\mathscr{S}_j + v \backslash \mathscr{S}_i$ where $d_t[S_i, \Psi'](x)$ was largest.

There is one quirk worth mentioning about these metrics: namely, the same Ψ can have $d_{\mathcal{V}} = 0$ and $d_t > 0$, but not vice versa. This is because each transition in $\mathcal{T}$ may not have consistent Γ's – i.e., the same process may produce different components depending on the transition. Thus Equation 3.29 is primarily useful as a starting point to finding any viable perturbation, where once $d_{\mathcal{V}} = 0$, one can switch to any d_t and proceed as described above.

3.6. Rule Derivations

The rule derivations for reversible organisations in empty (1,2)-environments described in Section 2.6 also extend to RealLife, and are in fact more powerful in this case, as we will see.

We will again need to compute 8 parameters $(b_0^L, b_0^U, b_1^L, b_0^U, s_0^L, s_1^U)$, where

$$b_0^L < b_0 \le b_0^U \tag{3.30}$$

$$b_1^L \le b_1 < b_1^U \tag{3.31}$$

$$s_0^L < s_0 \le s_0^U \tag{3.32}$$

$$s_1^L \le s_1 < s_1^U \tag{3.33}$$

$$0 < s_0 \le b_0 < b_1 \le s_1 \le 1. \tag{3.34}$$

Note that this last inequality is exactly the same as the general rule constraint given in Section 3.1. We then want to partition the processes of each transition in $t \in \mathcal{T}$ into four sets, corresponding to the process classes:

$$\mathcal{X}_t^b := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 0 \wedge \mathbf{s}_j(x) = 1\} \tag{3.35}$$

$$\mathcal{X}_t^{\neg b} := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 0 \wedge \mathbf{s}_j(x) = 0\} \tag{3.36}$$

$$\mathcal{X}_t^s := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 1 \wedge \mathbf{s}_j(x) = 1\} \tag{3.37}$$

$$\mathcal{X}_t^{\neg s} := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 1 \wedge \mathbf{s}_j(x) = 0\}. \tag{3.38}$$

We need to split the partitions per structure, since the calculations we will do become extremely cumbersome when expressed purely in terms of the graphon.

Now we compute the 8 parameters, assuming structure functions are defined so as to be 0 everywhere outside $\mathscr{S}$:

$$b_0^U := \min\left\{\inf\left\{\Psi(x) : x \in \mathcal{X}_t^b\right\}, t \in \mathcal{T}\right\} \tag{3.39}$$

$$b_0^L := \max\left\{\sup\left\{\Psi(x) < b_0^U : x \in \mathcal{X}_t^{\neg b}\right\}, t \in \mathcal{T}\right\} \tag{3.40}$$

$$b_1^L := \max\left\{\sup\left\{\Psi(x) : x \in \mathcal{X}_t^b\right\}, t \in \mathcal{T}\right\} \tag{3.41}$$

$$b_1^U := \min\left\{\inf\left\{\Psi(x) > b_1^L : x \in \mathcal{X}_t^{\neg b}\right\}, t \in \mathcal{T}\right\} \tag{3.42}$$

$$s_0^U := \min\{\inf\{\Psi(x) : x \in \mathcal{X}_t^s\}, t \in \mathcal{T}\} \tag{3.43}$$

$$s_0^L := \max\left\{\sup\left\{\Psi(x) < s_0^U : x \in \mathcal{X}_t^{\neg s}\right\}, t \in \mathcal{T}\right\} \tag{3.44}$$

$$s_1^L := \max\{\sup\{\Psi(x) : x \in \mathcal{X}_t^s\}, t \in \mathcal{T}\} \tag{3.45}$$

$$s_1^U := \min\left\{\inf\left\{\Psi(x) > s_1^L : x \in \mathcal{X}_t^{\neg s}\right\}, t \in \mathcal{T}\right\} \tag{3.46}$$

$$\tag{3.47}$$

Some of these values will not be defined when any of the sets $\mathcal{X}_t$ are empty for all t. These situations are handled as follows. All undefined values can be ignored in the inequalities above. If the parameters is needed to define another, the appropriate step depends on the parameter. If $\mathcal{X}_t^b$ is always empty, b_0^U should be ignored in the definition of b_0^L, and b_1^U should be ignored in its corresponding inequality. Otherwise, any undefined parameter used in the definition of another should be treated as if the extra condition were removed.

The logic of this is grounded in the following. First, $\mathcal{X}_t^{\neg b}$ must always contain a process that produces a 0-component due to being sufficiently *below* the b_0 boundary. This is always true since all viable configurations have ρ-wide boundaries; thus, in a vacuum, $\Psi = 0$ for all processes in $\partial\mathscr{S}$. This then means that there must always exist densities below b_0^L. Second, since Ψ is continuous, $\Psi > b_1$ only if $\mathcal{X}_t^b$ is non-empty: there must be an intermediate $\Psi \in [b_0, b_1]$ between those processes with $\Psi = 0$ and those with $\Psi > b_1$. Thus, b_0^L without the b_0^U term will only be defined over processes strictly below b_0 – the definition of b_0^L will never confuse densities above b_1 for those densities below. Finally, I assume that $\mathcal{X}_t^s$ is always non-empty.

The reader should recall that these conditions are precisely the same as those given in the LtL derivation. The main difference here is that the first two conditions are *necessary* in RealLife, but merely *sufficient* in LtL. This is what grounds my assertion that derivation is strictly stronger in RealLife. It is unclear whether the third condition ($\mathcal{X}^s$ always non-empty) is also necessary, but it certainly holds for any of the canonical examples of emergent individuals (blocks, blinkers, bugs with stomachs, etc.). Thus for any organisation satisfying this last condition, this rule derivation procedure is *exhaustive*.

There is one last oddity to note, however. If a set of parameters gives rise to an inequality of the form $a < \{b_0, b_1, s_0, s_1\} \leq a$, we can set the parameter to a without trouble. The reason I use the non-inclusive bounds by default is because: (i) it is consistent with the LtL definitions; and (ii) it guarantees that some of the nice topological features over $\mathcal{U}$ described above can be preserved (e.g., continuity of ξ). But both of these are not essential qualities for our purposes.

Example 5: The Block. The block defined in Example 1 is partitioned into two sets: $\mathcal{X}^s = \odot_{\rho/2}$ and $\mathcal{X}^{\neg b} = \odot_{3\rho/2} \setminus \odot_{\rho/2}$. This gives:

$$b_0^L = \max\{\sup\{\Psi(x) : x \in \mathcal{X}^{\neg b}\}\} = \frac{\lambda\left[\odot_{\rho/2}\right]}{\lambda[\mathcal{N}]} = \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s]$$

$$s_1^L = \max\{\sup\{\Psi(x) : x \in \mathcal{X}^s\}\} = \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s]$$

$$s_0^U = \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s].$$

This gives the inequalities:

$$b_0 > \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s]$$

$$s_0 \leq \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s]$$

$$s_1 \geq \lambda[\mathcal{N}]^{-1}\lambda[\mathcal{X}^s]$$

$$0 < s_0 \leq b_0 < b_1 \leq s_1 \leq 1.$$

Any RealLife rule that satisfies these will support the block organisation in an empty environment.

To demonstrate the power of this method further, we can rewrite the inequalities in Example 5 – assuming the general rule constraint – as $s_0 \leq \lambda[\mathscr{N}]^{-1}\lambda[\mathcal{X}^s] < b_0$ (we drop s_1 since it is implied by $b_0 < b_1 \leq s_1$). Hence,

$$\frac{\lambda\left[\bigodot_{\rho/2}\right]}{\lambda[\mathscr{N}]} \in [s_0, b_0). \tag{3.48}$$

It turns out that this is equivalent to the result Pivato proved for the existence still-lifes (Pivato, 2007, Proposition 3.3). We thus have strong evidence that this derivation is exhaustive.

Example 6: The Blinker. The blinker graphon defined in Example 2 has all $\mathcal{X}_t^*$ non empty, resulting in the following parameters:

$$\begin{array}{lll} b_0^U = \frac{3}{16} & b_0^L = \frac{3}{16} & b_1^L = \frac{1}{2} \\ s_0^U = \frac{3}{8} & s_0^L = \frac{3}{8} & s_1^L = \frac{1}{2} \end{array}$$

This in turn gives rise to the family of rules that support the existence of a $\rho \times 2\rho$ blinker:

$$\left(\rho, b_0 = \frac{3}{16}, b_1 \geq \frac{1}{2}, s_0 = \frac{3}{8}, s_1 \geq \frac{1}{2}\right). \tag{3.49}$$

If Example 6 can be proven as the necessary and sufficient conditions for the existence of blinkers, then the autopoietic theory presented in this chapter offers an alternative possibility to scaling emergent patterns from CAs into EAs and larger CAs that has not been used in the literature on LtL and RealLife.

Chapter 4

Discussion

Autopoiesis, as a theory of biological identity, has the potential to transform how we conceive of biological phenomena as whole. It offers a perspective that distinguishes between the possibilities for and the realisations of an organism. This in turn has implications for how we ought to understand cognition as a matter of structural coupling constrained by organisation; and for how we understand viability as the limits of structural possibility determined by the organisation.

Cellular automata can serve as useful model universes in which ideas in theoretical biology concerning emergent individuals can be tested by developing mathematical theory in simple, concrete cases. Thus, a formalisation of autopoiesis in these universes is a good first step toward a more mature theory of biological identity.

In the preceding chapters, I sought to extend the autopoietic theory developed by Beer for the GoL universe (Beer, 2004, 2014, 2015, 2020b; Beer, McShaffrey, & Gaul, 2024) to a larger neighbourhood generalisation, LtL, and its continuum limit, RealLife. I was successful in this endeavour: I showed that the theory can be extended to larger neighbourhood individuals in LtL by computing their process dependency graphs and deriving from this their cognitive domains and intrinsic viability constraints. Moreover, the way I defined these graphs is different than has appeared before, in order to allow for scaling to RealLife, and this definition introduced new complications. I also introduced the density formulation of the viability constraint as a potentially more useful representation. Finally, I developed a novel method for deriving rules from process dependency graphs such that those rules allow the realisation of a reversible organisation in empty universes.

In RealLife, I defined the autopoietic organisation of emergent individuals using graphons, and showed that cognitive domains and viability constraints could be derived just the same as in LtL. I also presented evidence that graphons are the appropriate limit object of sequences of process dependency graphs of increasing neighbourhood size, thus establishing a continuity between the discrete and continuous theory. Finally, I extended the rule derivation method to RealLife and showed that it is stronger in the continuous case.

4.1. Autopoiesis

The interpretation of autopoiesis I have formalised emphasises a number of key concepts. First, both space and time are indispensable for autopoiesis to hold the theoretical power it claims. Without defining a reaction in LtL as a spatial arrangement of reactants, or without the inherent spatial and temporal ordering of enabling relations, the derivation of the cognitive domain would be impossible. The interactions a system can engage in cannot be described without accounting for the spatiotemporal structure of the system and how this structure creates an environment distinct from that system.

Second, the physical boundary is indispensable. It is the only way for the system to distinguish itself from the environment and thus establish a distinct identity. For example, consider a universe of all 1-components. If we ignored the boundary and only defined an individual by its 1-components, we could identity every individual at every point in space simultaneously. Clearly, it does not make sense to talk about the interactions such an individual can engage in with its environment when it is not meaningfully distinct from that environment. It is therefore imperative that any formalisation of autopoiesis include a physical boundary such that a system can be usefully described as a cognitive system acting in space as a distinct unity.

Third, the observer plays a significant role defining the organisation of an individual. As I discussed in Section 2.2, the organisation-structure duality implies an insufficiency of empirical observations in determining the organisation. If we again consider the glider, an observation of any one of its reversible subgraphs is insufficient to determine the full space of its interaction. That is, such an observation could only determine a subset of the possible concatenations of processes implied by the full organisation. It is up to the observer to introduce additional criteria to expand from limited observations, including considerations of symmetry, basins of attraction, etc. However, once the organisation is defined, it fully determines the space of its possible realisations. To be clear, the observer is still involved in this determination to a certain extent, since this space only arises when the observer considers the organisation in a particular context (e.g., its realisation in space). But this context is minimal, as it is simply a matter of the organisation reducing the full space of possibilities (every possible configuration of components) to the subset of those that realise the organisation. And importantly, all information used in this reduction derives entirely from the organisation through the implicit spatial embedding of its processes.

The observer-dependence of the organisation hints at a deeper aspect of autopoietic theory (or at least in my interpretation): there is no metaphysical status granted to autopoietic systems, or to any system whatever. This is a consequence of the enactive understanding of experience as the epistemologically primary phenomenon. Thus, any notion of an external world of objects is necessarily relative to an observer. More generally, all concepts are interpretations of experience. Organisation is no different: the "correct" organisation to describe an organism is whatever is most useful and satisfies the scientific norms of the observer-community. This is not to say that the relation between organism and organisation is arbitrary, but that the way in which the organism constrains our description of it is decided by the observer-community. The GoL block organisation is clearly not a very good model of *E. coli* because it is not very effective at doing what we would want such a model to do (predict behaviour, viability, etc.).[1]

4.1.1. Significance of the Formalism

It is worth further interrogating the significance of the degeneracies that arose in both LtL and RealLife: there were processes that neither produced any components of the system, nor had their center-components produced by any other process in the organisation, and this lead to a failure of the cognitive domain derivation. This degeneracy, however, was limited to reversible blinkers, and is therefore not a generic feature of the theory. It is also dependent on some peculiarities of the chemistry, such as the immobility of components and the immediate consumption of products, i.e., a product cannot move to become a reactant in another process, or persist to become a reactant at a later time. It is thus unclear what significance this degeneracy could have beyond LtL and RealLife. This problem can also be avoided entirely, at least in the discrete case, by using the vertex-centric formulation of process dependency graphs, since every relevant component would then be specified by at least one process in the organisation.

The similarity between LtL and RealLife reduces the novelty of the latter with respect to our understanding of autopoietic theory. That is, RealLife introduces nothing fundamentally new to address most of the abstract questions of autopoiesis. However, RealLife is very informative of what a more sophisticated theory of autopoiesis might look like. For instance, an organisation in continuous-time continuous-space universes would almost certainly have infinitely many processes; RealLife shows how one might handle that problem while preserving representation independence. It also introduces measure equivalence in almost every aspect of the theory, which is likely to generalise to more realistic models (and especially continuous-time models). Furthermore, the convergence of LtL to RealLife

[1] Of course, the norms we use to evaluate a model are themselves maleable, and it is worth interrogating whether they are sufficient to determine, say, what a proper theory of autopoiesis should look like. For instance, we might want to consider whether a model is corroborated by phenomenological evidence, and this might help to constrain the space of possible organisations that describe a system.

demonstrates the possibility of a distinction between exact and approximate theories, and how this might make a theory of autopoiesis both rigorous and practical. Finally, it gives at least one example beyond the discrete case of what sort of calculations could be made with autopoietic systems, and how to carry them out.

Most basically, this extension of autopoietic theory to RealLife suggests the possibility of further extensions and establishes a blueprint for how to achieve them.

4.2. COGNITION

The autopoietic perspective on cognition emphasises the co-dependence between an individual and its environment in determining behaviour. Organisation does not dictate behaviour, but merely the possibilities thereof – it is only when coupled to an environment does cognition really happen.

This perspective suggests a continuity in how we ought to understand cognition in more complicated systems. That is, the basic ideas of structural coupling constrained by organisation is universal to all behaving systems of interest to cognitive scientists. All that changes from system to system, and from situation to situation, is the particular form this coupling takes. Already we can see this in LtL by comparing a block to the glider. A block can only engage in one possible interaction, or else its organisation disintegrates. The glider, on the other hand, always has many interactive possibilities at any given moment. Depending on the dynamics of the environment, it can undergo different sequences of structural transitions, which in turn reciprocally alter the dynamics of the environment (Beer, 2020a).

This idea is also reflected in embodied and dynamical approaches to cognitive science, and in particular the minimally cognitive behaviour research programme (Beer, 1995a, 1997, 2008; Favela, 2020). Here, cognitive behaviour is understood as structural coupling between an agent, represented as a nonautonomous dynamical system, and its environment, which are together analysed as an integrated whole using the tools of Dynamical

Systems Theory (Beer, 1995b; Garfinkel et al., 2017; Strogatz, 2024). The connection between this and the autopoietic theory presented in the preceding is, at least conceptually, the derivation of the cognitive domain. That is, an interaction graph can be interpreted as a nonautonomous dynamical system where the state space is the set of unique structures, and the possible evolutions of the system are described by the edges. A theory of autopoiesis then serves as a foundation to a theory of cognitive behaviour, where the latter is concerned with the structural coupling of a given system, and the former is concerned with the emergence of that system and how its organisation determines the possibilities for structural coupling.

The structure-organisation distinction also allows us to talk both about behaviour as it is realised, and about the perspective of the individual as it behaves. In the context of LtL and RealLife, this amounts to a distinction between the interaction graph and a path through it. The former describes all the ways that the environment is relevant to the individual by partitioning the space of perturbations according to how the individual responds to them. In some sense, then, every perturbation "looks a certain way" to the individual since it will respond to every perturbation of the same class in exactly the same manner. For instance, the $\rho = 5$ blinker presented in Chapter 2 has an interaction graph with three distinct perturbation classes: `GRAY`, `ORANGE`, and `GREEN`. Every perturbation that does not destroy the blinker must fall into one of these classes. We can therefore recast structural coupling in terms of these classes, and thus describe the world of the blinker as it unfolds from its own perspective (Beer, 2014).

In this way, the interaction graph can be interpreted as a formal description of the *Umwelt* of an individual (von Uexküll, 1934/1992) – what the world is like from its own perspective. This points at a broader theme: experience can be understood in terms of the distinctions a system makes (Gaul & Izquierdo, 2025). By mapping out these distinctions in the cognitive domain, the observer gains the ability to simultaneously consider and mediate between the environment an organism is situated in, and the world that the organism brings forth through structural coupling (Varela, 1997; Varela et al., 2017).

4.3. VIABILITY

The core difference between the formulation of viability presented in Chapters 2 and 3 (and in Beer, McShaffrey, and Gaul, 2024) and the formulation of viability following Ashby (1960) amounts to the distinction between *intrinsic* and *extrinsic* viability constraints. Extrinsic constraints are imposed on a pregiven system by the observer, such that a system is considered non-viable only insofar as its state exceeds the bounds set by the observer – the system need not even be capable of disintegration. In contrast, an intrinsic constraint is derived from the organisation of a (precarious) system and is thus only dependent on the observer insofar as its identity as an individual is so dependent.

This core distinction gives rise to many interesting differences between the formulations of viability. For one, the concept of essential variables is complicated by the fact that an autopoietic system in LtL or RealLife may depend on components in the environment to constitute itself at the next step – the essential variables of the system are not restricted to just the components and processes contained within its boundary. Moreover, what variables are essential will change over time, according to the structure of the system. Another difference arises from how the boundary of the constraint is defined. In the intrinsic constraint, the boundary is defined as the configurations that contain a realisation of the organisation and a terminal perturbation to it; in an extrinsic constraint, it is defined simply as the boundary of the states chosen by the observer. Extrinsic constraints therefore permit states at the boundary to be non-terminal (an important feature in viability decomposition: McShaffrey and Beer, 2023). The basis of this difference lies in what is constrained in either case: an extrinsic constraint sets bounds on the states (structures) of a system, whereas an intrinsic constraint sets bounds on both structures *and* structural transitions – an intrinsic constraint is a constraint on behaviour. Thus, while it may make sense to talk about how changes to the dynamics of a system change its behaviour at the boundary of viability in the extrinsic case, such questions are trivial in the intrinsic case: the behaviour is always the same (i.e., terminal).

This is not to say that intrinsic constraints are necessarily superior, nor that they capture everything that an extrinsic constraint can. It would therefore be interesting to see what could be done to bring the two approaches together. One option could be to apply viability decomposition to an intrinsic constraint by adding a topology derived from a system's possible trajectories of structural coupling (Beer, 2020a). With this, trajectories that never reach the boundary are classified as *asymptotically viable*, while those that eventually terminate are *transiently viable* (McShaffrey & Beer, 2022; McShaffrey & Beer, 2023). Note that this would still result in many key differences. For example, a trajectory couldn't approach the boundary and not be terminal since the topology is defined not by the space of structures, but by structural coupling.

4.4. Adaptivity and Sense-making

The intrinsic formulation of viability immediately present problems for the operationalisation of adaptivity (Di Paolo, 2005, 2009). Specifically, Di Paolo (2005, p. 438) defines adaptivity as

> A system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability,
>
> 1. Tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence,
> 2. Tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with outward velocity.

That is, a system is adaptive if it can sense the limits of its viability and accordingly alter its behaviour so as to avoid its internal state approaching those limits too closely. The problematic points here are "sensing the limits of viability" and "accordingly alter its behaviour."

With respect to the first, a system may not have access to the information necessary to determine the limits of viability. For example, any autopoietic system in LtL – unless it is a still-life – is dependent on components in its 2-environment in every structural transition. Such a system is therefore incapable of determining whether it is in the boundary of its constraint (a terminal environment), or in the interior. With respect to the second problematic point, an intrinsic constraint necessarily captures all possible behaviour of a system, as implied by its organisation. Therefore, even if we grant that a system can determine whether it is in the boundary of the viability constraint, or even its distance from that boundary, it would be impossible for it to change that fact, since any such change would already be accounted for. In other words, an adaptive system does not change its trajectory with respect to the limits of viability, but is in an adaptive trajectory from the start.

The source of these problems lies in the implicit decomposition of an autopoietic system into "regulator," on the one hand, and "the system," on the other. It only makes sense to talk about a system changing its trajectory with respect to the viability constraint if some of its components are ignored when describing the constraint. But what motivates such a decomposition? If it is simply a matter of explaining the role some mechanism plays in the behaviour of the system, then it is misleading to talk about the constraint of the subsystem as *the* constraint. This is not to say that such a decomposition could not be useful, but that our explanatory needs as observers are independent of the organisation and should therefore not determine its viability.

The consequence of all this is that adaptivity only makes sense as a particular class of behaviour: when a system avoids situations that we, as observers, determine it would otherwise lose viability in. This deflation of adaptivity eliminates the particular operationalisation of "graded norms" that Di Paolo (2005) sought. There is no sense in which a system can be close or far from the limits of its viability without such distance corresponding precisely to the actual lifetime of the system. Moreover, it is not enough to say that a system can use proxies to the viability constraint, and that this is sufficient to recover a graded normativity intrinsic to the system. Again, the full viability constraint is something that only an observer has access to, and any connection made between a component of the system and its viability constraint is strictly mediated by the observer – our explanatory needs should not be confused with the actual operation of the system.

This is not to say that one cannot talk about situations being better or worse for a system, nor that a system cannot make a distinction between the two, but that such a distinction is strictly a behavioural phenomenon, independent of the context we put it in. In other words, notions of "good" or "bad," insofar as they are scientifically useful, are proper to the cognitive domain of an individual and its structural coupling.

If we can no longer use the stronger notion of adaptivity, what of its supposed implications for normativity, agency, teleology, intentionality, sense-making, and so on? Properly addressing these issues requires a deeper engagement with the phenomenological literature surrounding adaptivity and enaction (Di Paolo, 2005; Jonas, 2001; Merleau-Ponty, 1942/1963, 1945/2010; Thompson, 2007; Varela et al., 2017; Weber & Varela, 2002), and so I do not pursue this here. However, I will offer a sketch of what an alterative grounding of some of these concepts might look like.

To begin, it will be useful to layout in more concrete terms what adaptivity hoped to achieve. It's primary aim was to establish a bridge from autopoiesis to *sense-making*, as proposed by Weber and Varela (2002) (Di Paolo, 2005, p. 430). Sense-making is understood to be the instauration of a meaningful perspective on the world, and meaningful in a *normative* sense: things in the world are good or bad with respect to the continued autopoiesis of the system. Thus, this is supposed to explain fatigue and illness as states that are bad for the system, with respect to its intrinsic norm (Di Paolo, 2005, p. 440). Adaptivity is also meant to distinguish between *systems* that are acted upon by the environment, and *agents* that have an asymmetric interactive relation to their surroundings by modulating those very interactions. Lastly, a system's intrinsic normativity is taken to imply an intentionality whereby the system is always directed toward some future beneficial situation (or away from a deleterious one).

Taking these in order, I will start with sense-making. The establishment of a perspective follows simply from autopoiesis: the system partitions all perturbations by how it responds to them, thus picking out what is relevant to it. The question for sense-making then becomes whether such a perspective is necessarily *meaningful*. If we broaden our sense of what makes something meaningful beyond strictly good or bad with respect to autopoiesis, then any perspective is inherently meaningful. Specifically, a perturbation is meaningful in precisely the way it is situated within the cognitive domain of the system. For instance, any `BLUE` perturbation on the glider is meaningful for it *as a* `BLUE` *perturbation*. This means that the sense in which something is meaningful need not be intuitive to the observer. Practically, we can only understand the meaning something has for a system by considering that thing in the context of the system's way of life.[2] Thus, although being maladaptive is properly a concern of the observer (by situating behaviour in the context of viability outcomes), it can still makes sense to talk about a situation as bad *from the perspective of the system*. For example, *E. coli* understands low nutrient concentrations as bad *because it avoids those situations*.

An important point about this notion of a meaningful perspective is that it is non-teleological.[3] "Good" and "bad" are abstractions from the behaviour of a system, and need not correspond to any objective feature of it. This is not to say that value is somehow "not real," but that its reality is grounded in the way of life brought forth by a system in structural coupling. For example, we say injury is bad, not because there is some inherent feature of the experience of pain that says "this is bad," but because we tend to avoid injury (Wittgenstein, 1953/2009). Note that the cognitive domain is still essential here. Without it, a behavioural mimic would satisfy this notion of value just as well. Avoidance only makes sense as a *possible* action given by the cognitive domain.

This same argument can be applied to illness and fatigue: they are bad because we avoid or seek to remedy them. To make this clearer, consider the experience of color, a key example in the original statement of enaction (Varela et al., 2017). According to enaction,

[2] This is analogous to Wittgenstein's 1953/2009 idea that the meaning of a word is revealed by the use to which it is put in the context of a *language-game*: the language used in a particular pragmatic situation.

[3] It is worth mentioning that Ashby (1960) was emphatic that his notion of *ultrastability*, similar to adaptivity, was non-teleological.

the experience of color is something we bring forth through our responses to variations in frequency of light (Varela et al., 2017, chap. 8). Importantly, nothing here requires that this experience correspond to something within the system. We must therefore ask why it is necessary for the experience of value to be operationalised as distance from a viability boundary (something apparently intrinsic to the system), but color can be explained as simply emergent from structural coupling. This hints at a problem with Di Paolo's (2005, 2009) naturalisation of normativity in being implicitly representational, which is in direct tension with the tenets of enaction that it is supposed to support (Barrett, 2017).

The concept of asymmetric interaction between an agent and its environment is only explained by adaptivity through its dependence on the implicit decomposition discussed above. If we can only consider whole systems in structural coupling with their environment, there is no room for any added regulation of interaction since this would already be a part of the ongoing structural coupling. However, there is something to the intuition of asymmetry here: there is a difference between a system passively absorbing whatever nutrients are around it, and a system actively seeking nutrients out. Perhaps this intuition can be captured in the following way: a system is an asymmetric interactive relation with its environment to the extent that the modes of structural coupling it engages in are sensitive to changes in internal state. For example, an *E. coli* can change between running and tumbling due simply to its internal state (Sourjik & Wingreen, 2012). In contrast, an immotile bacterium would be incapable of such a change, irrespective of its internal state. Clearly, this conception of asymmetry depends on a more rigorous notion of what constitutes a "mode of structural coupling." But if one can make this notion rigorous, then it should be possible to apply this to dynamical models and compare the extent to which different systems have interactive asymmetries.

Finally, I come to intentionality. If we are concerned only with intentionality as the directness toward a virtual future, it is unclear what this means without imposing a teleology. However, there is still a sense in which an autopoietic system is directed in its experience through the distinctions it makes in its cognitive domain. That is, an autopoietic system is always oriented toward its future by acting upon the distinctions it makes in the world. In fact, an autopoietic system must, by definition of the cognitive domain, act upon its distinctions. The question then becomes whether this notion of intentionality is sufficient to explain experience. This I am unsure of, and, again, a proper treatment of this issue requires a deeper engagement with the phenomenological tradition that takes us beyond my immediate concerns here. What I provide here is simply a sketch, not a fully developed account of the connection between autopoiesis and enaction. Moreover, such an account cannot be properly developed simply through philosophical discussion, but requires further theoretical development in more realistic models.

It is now worth re-interrogating whether autopoiesis, on its own, is sufficient for what adaptivity is supposed to explain. I think it is sufficient for sense-making as a meaningful perspective on the world, even if this meaning is not obvious to us. However, it is not sufficient for interactive asymmetry. Whether it is sufficient for the valence in the sense of good/bad depends on what one means by those terms. My intuition is that nothing can be said about this *in general*, as what "good" or "bad" mean will always be relative to a particular system and the sort of life it lives. Again, good/bad need not be an objective feature of living systems, but rather something they bring forth through their emergent subjectivity.

4.5. Future Directions

There are many possible extensions that can be explored in LtL and RealLife. Most obvious would be to generalise Beer's work on structural coupling and the origins of autopoiesis in GoL (Beer, 2020a, 2020c). Another option could be to explore different sets of structures besides those that occur in a vacuum (for example, the intermediate block configuration in Figure 2.7). One could also make some the results I presented more rigorous by, for instance, proving the convergence of process dependency graphs, or proving the sufficiency of the rule derivation procedures. There are also some of the problems I left unresolved that may yet have solutions: degenerate organisations, assumption of vacuum perturbations in rule derivations, exhaustive LtL rule derivations, multigraphs in RealLife, topological assumption in non-reversible RealLife organisations, etc.

However, it may be more productive to move on to different model universes with more realistic features. As noted above with respect to the degeneracies of process dependency graphs, there are some peculiar features of the LtL and RealLife chemistries: the immobility of components and the immediate consumption of products. Relatedly, there is also no diffusion in these chemistries. There are thus a number of possible extensions that include some of these phenomena, including Lattice Gas Automata and Lattice Boltzmann models (Wolf-Gladrow, 2000). Moreover, these have a proven continuum limit in Navier-Stokes equations (Frisch et al., 1986; Krüger et al., 2017). These models also have the advantage of being common tools in physics with many guidelines and rigorous results for them.

More broadly, a mathematical theory of autopoiesis has many potential applications. There is obvious biomedical relevance in being able to predict what perturbations will most effectively destroy a cell. Similarly, being able to find perturbations to manipulate organisms is of relevance to bioengineering. Molecular biology would also benefit from an autopoietic perspective by revealing the significance particular components have for the persistence of an organism's organisation. Ethology and cognitive science could also benefit from a more embodied dynamical perspective on agent-environment interaction, as has been argued extensively elsewhere (Beer, 1995a, 2023; Favela, 2020; Thelen & Smith, 1994; Varela et al., 2017).

***

An autopoietic perspective on life and cognition has the potential to transform how we think about both the scope and depth of biological phenomena as a whole. To fully realise this potential, rigorous theories need to be developed in iteratively complexified models, stepping toward a mature theory of autopoiesis in our own universe. My work here has merely been a small step in this direction.

Appendix A

Definitions and Algorithms

Throughout the main text of this thesis, I have left many statements unsubstantiated, or else imprecise, for sake of clarity. The following seeks to remedy this by providing formal definitions and arguments. In other words, this appendix is for the mathematician and pedant.

Section A.1 describes the theory for LtL, and Section A.2 RealLife. Section A.3 describes the connection between the two that shows the LtL theory to be an approximation of the RealLife theory. Between the first two sections, I use much of the same notation for slightly different objects that nonetheless serve the same role. Some of these objects will need to be distinguished more carefully in Section A.3.

There are a few other abuses of notation to note. First, I will often add or subtract sets and scalars to mean applying that arithmetic to every element of the set: $\mathscr{A} + v = \{a + v : a \in \mathscr{A}\}$. Second, if a function is, strictly speaking, set-valued, but I have some guarantee that the elements of that set (or their images) are all the same, then I will often treat the function as if its value was one of those elements, since how one choses an element would then not matter. Third, I will usually suppress dependence on the parameters $(\rho, s_0, b_0, b_1, s_1)$ and p to simplify notation.

Finally, Section A.1 describes LtL as a proper theory in its own right of a discrete universe. It is not until Section A.3 that I will treat it as an approximate theory of autopoiesis in RealLife.

A.1. Larger than Life

Let $\mathscr{M} := \{0, 1\}$ and let be the set of possible cell states. A *universe state function* is a function $\mathbf{u} : \mathbb{Z}^D \to \mathscr{M}$. We denote the space of all universe states $\mathcal{U} := \mathscr{M}^{\mathbb{Z}^D}$. Let the *neighbourhood* be a set $\mathscr{N} := \{x : \|x\|_p \leq \rho\}$ centred at the origin, where $\rho \in \mathbb{N}$ is a finite neighbourhood radius.

Let $\kappa(x) := |\mathscr{N}|^{-1} \mathbb{1}_{\mathscr{N}}(x)$ be a convolution kernel. The *density function* $\Psi : \mathbb{Z}^D \to [0, 1]$ is defined as a convolution, giving the average value about a point x:

$$\Psi(x) := (\kappa * \mathbf{u})(x) = |\mathscr{N}|^{-1} \sum_{y \in \mathscr{N}} \mathbf{u}(x - y)\, \lambda[dy] \tag{A.1}$$

A *D-dimensional Larger than Life cellular automaton* (Evans, 2001, 2003) is a functional $\xi : \mathcal{U} \to \mathcal{U}$, defined

$$\xi[\mathbf{u}](x) := \mathbf{u}(x) \mathbb{1}_{[s_0, s_1]}(\Psi(x)) + (1 - \mathbf{u}(x))\, \mathbb{1}_{[b_0, b_1]}(\Psi(x)), \tag{A.2}$$

where $0 \leq s_0 \leq b_0 \leq b_1 \leq s_1 \leq 1$. Thus a LtL rule can be fully specified by a 5-tuple $(\rho, s_0, b_0, b_1, s_1)$ and a choice of p-norm.

It will be useful to define a few additional sets and functions. Let $\sigma^v[\mathbf{u}](x) := \mathbf{u}(x - v)$ be a functional that translates state functions by some $v \in \mathbb{Z}^D$. Let $\mathcal{E}(\mathscr{A}) := (\bigcup_{x \in \mathscr{A}} \mathscr{N} + x) \backslash \mathscr{S}$ give the 1-environment of $\mathscr{A} \subset \mathbb{Z}^D$. We get the n-environment by recursion: $\mathcal{E}^n$. With a slight abuse of notation, we will similarly use $\mathcal{E}^n(\mathbf{s})$ to denote the set of functions over $\mathcal{E}^n(\mathscr{A})$ that are equal over $\mathscr{A}$.

A.1.1. Autopoiesis

Definition A.1 (Processes). *A* process *in Larger than Life is a tuple* $\psi = (\phi, \mathbf{p})$ *where* $\phi \subseteq \mathscr{N} + v$ *for some finite* $v \in \mathbb{Z}^2$*, and* $\mathbf{p} : \phi \to \mathscr{M}$ *is well-defined over* ϕ*. Two processes,* $(\phi_i, \mathbf{p}_i)$ *and* $(\phi_j, \mathbf{p}_j)$*, are* chemically equivalent *if, when translated to the origin (*$\phi \mapsto \phi - v$*):*

1. **(Equivalence of space)** $\phi_i = \phi_j$,

2. **(Equivalence of state)** $\mathrm{supp}(\mathbf{p}_i) \cap \phi_i = \mathrm{supp}(\mathbf{p}_j) \cap \phi_j$.

3. **(Equivalence of type)** $\mathbf{p}_i(\vec{0}) = \mathbf{p}_j(\vec{0})$.

Definition A.2 (Structures). *Let* $\mathscr{S} \subset \mathbb{Z}^D$ *be a finite, connected set and let* $\mathbf{s} \in \mathcal{U}$ *be a function well-defined over* $\mathscr{S}$*. Then* $(\mathscr{S}, \mathbf{s})$ *is a valid* structure *if* $|\mathrm{supp}(\mathbf{s}) \cap \mathscr{S}| > 0$ *and* $\mathscr{S} = \bigcup_{x \in \mathrm{supp}(\mathbf{s})} \mathscr{N} + x$*. Two structures* $(\mathscr{S}_i, \mathbf{s}_i)$ *and* $(\mathscr{S}_j, \mathbf{s}_j)$ *are equivalent if they are translationally symmetric, that is, if there exists some* $v \in \mathbb{Z}^D$ *such that* $\mathscr{S}_i + v = \mathscr{S}_j$ *and* $\sigma^v \circ \mathbf{s}_i = \mathbf{s}_j$*.*

Definition A.3 (Process Dependency Graph). *Let* V *be an arbitrary finite set and* $K := \mathscr{N} \times \mathscr{M}^2$ *a set of edge labels, where* $k = (r, c_0, c_1) \in K$*. Let* k_r*,* k_{c_0}*, and* k_{c_1} *give the component of the edge label corresponding to the subscript. Let* E *be a set of triples* (i,j,k) *where* $i,j \in V$ *and* $k \in K$*. Finally, let* $\mathfrak{s} : E \to V$ *give the* i *component of the edge and let* $\mathfrak{t} : E \to V$ *give the* j *component. Then the multidigraph* $G = (V, E, \mathfrak{s}, \mathfrak{t})$ *is a valid* process dependency graph *if the following conditions are satisfied, where* $\psi_i = (\phi_i, \mathbf{p}_i)$ *is the process associated with the vertex* i*:*

1. **(Full Enablement)** *For every* $i \in V$*, there exists a* $(\cdot, i, k) \in E$ *for each* $k_r \in \phi_i$*.*

2. **(Consistent Enablement)** *For every* $i \in V$*,* $\mathbf{p}_i(x \in \phi_i) = c$ *iff* $c = k_{c_1}$ *for all* $(\cdot, i, k) \in E$ *where* $k_r = x$*, or* $c = k_{c_0}$ *for all* $(i, \cdot, k)$ *where* $k_r = x = \vec{0}$*. And for every* $i \in V$*,* c_0 *is constant for all* $(i, \cdot, (\cdot, c_0, \cdot)) \in E$

3. **(Connectivity)** *G is weakly connected.*

G is reversible *when, for every $i \in V$ and $r \in \mathscr{N}$ such that $\exists\ (\cdot, i,\ (r, \cdot, \cdot)) \in E$, there is exactly one such edge.*

G is closed *when, for every $i \in V$, there exists an $(i, \cdot, \cdot) \in E$ and an $(\cdot, i, \cdot) \in E$.*

G is degenerate *if there exists some $i \in V$ such that there is no $(i, \cdot, \cdot) \in E$ and no $(\cdot, i, (\vec{0}, \cdot, \cdot)) \in E$. Any such i is corresponds to a degenerate process with respect to the organisation.*

Processes are represented implicitly here, but they can be extracted from the edges. Let the *center-component* of a process $i \in V$ be the unique c_0 such that $(i, \cdot, (\cdot, c_0, \cdot)) \in E$. The state c of the $r \in \mathscr{N}$ component of ψ_i is given by any edge $(i, \cdot, (r, \cdot, c)) \in E$.

Definition A.4 (Autopoiesis)**.** *A process dependency graph G (Definition A.3) is autopoietic if there exists a partition $\{O_i\}$ of V such that every O_i satisfies the following conditions:*

1. *There exists a bijection $f_i : O_i \to \mathscr{S}_i$ such that $(\mathscr{S}_i, \mathbf{p}_{f_i^{-1}}(\vec{0}))$ is a valid structure.*

2. *$(\mathscr{S}_i, \mathbf{s}_i)$ can be expressed as the union of translated processes in O_i.*

3. *There exists O_j such that for some $\mathbf{u} \in \mathcal{E}(\mathbf{s}_i)$ and a LtL rule, $\xi[\mathbf{u}]\big|_{\mathscr{S}_j + v} = \sigma^v \circ \mathbf{s}_j$, where $v = f_j(x_j) - f_i(x_i)$ such that $x_i \in O_i$, $x_j \in O_j$, and $\exists (x_i, x_j, k) \in E$ with $k_r = \vec{0}$. Let $\mathscr{D}_{ij} = \mathscr{S}_i \cap (\mathscr{S}_j + v)$. Then for every $f_i(x_i) \in \mathscr{D}_{ij}$, $\exists (x_i, x_j, k) \in E$ with $k_r = f_j(x_j) - f_i(x_i) - v$ for all x_j such that $f_j(x_j) + v \in (\mathscr{N} + f_i(x_i)) \cap (\mathscr{S}_j + v)$.*

4. *There exists O_j such that (3) is satisfied where O_i and O_j are swapped.*

The network as a whole must also be such that all $(x_i, x_j, k) \in E$ satisfy $f_i(x_i) \in \mathscr{D}_{ij}$ and $k_r = f_j(x_j) - f_i(x_i) - v_{ij}$ for some O_i, O_j valid according to condition (3).

According to this definition, then, an autopoietic process dependency graph implies a closed network of structures and every enabling relation participates in some transition between structures. Note that the graph itself need not be *completely* closed. It is possible to prune from the network those processes that do not enable any other process and still recover complete structures where they are implicit (assuming the graph is not degenerate).

However, this would require a more complicated statement that combines conditions (1) and (2), or else replace conditions (1)-(4) with “must derive a closed network of structure transitions” using the procedure outlined in Section A.1.2. As such, I will use the simpler definition of autopoiesis.

We can guarantee that only boundary processes are degenerate in autopoietic organisations, due to the light cone constraint (condition (3) in Definition A.4). This implies that, in any valid transition, the 1-environment cannot produce any 1-components that constitute the system, since then the boundary of the subsequent structure would exceed the light cone of the initial structure, which is not allowed. Thus, every 1-component must be produced directly by the processes realising the initial structure, meaning the center-component of the corresponding process in the subsequent structure is always directly enabled, i.e., it is not degenerate.

A.1.2. The Cognitive Domain

First, we must define the symmetries of a structure as a dihedral group of degree 4 and order 8, D_8, when the dimension of the LtL universe is $D = 2$. This group can be generated by two functionals representing rotations, $\mathbf{r}$, and reflections (flips), $\mathbf{f}$. To define these, I assum that structures are always *normalised*. Let $(\tilde{\mathscr{S}}, \tilde{\mathbf{s}})$ denote the normalised form of a structure: $\tilde{\mathscr{S}} := \mathscr{S} - v$ and $\tilde{\mathbf{s}} := \sigma^{-v} \circ \mathbf{s}$, where

$$v = \Big\langle \min\{x : \exists y \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\} - \frac{\max\{x : \exists y \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\} - \min\{x : \exists y \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\}}{2},$$
$$\min\{y : \exists x \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\} - \frac{\max\{y : \exists x \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\} - \min\{y : \exists x \text{ s.t. } \langle x,y\rangle \in \mathscr{S}\}}{2}\Big\rangle.$$

Then

$$\mathbf{f}[\tilde{\mathbf{s}}](x) := \mathbf{s}(\langle -x_0, x_1\rangle) \tag{A.3}$$

$$\mathbf{r}[\tilde{\mathbf{s}}](x) := \mathbf{s}\left(\left\langle \|x\|_2 \cos\left(\frac{\pi}{2} + \tan^{-1}\frac{x_1}{x_0}\right), \|x\|_2 \sin\left(\frac{\pi}{2} + \tan^{-1}\frac{x_1}{x_0}\right)\right\rangle\right) \tag{A.4}$$

generates the group $\mathfrak{R} = \langle\{\mathbf{f}, \mathbf{s}\}\rangle$, which is isomorphic to D_8. Thus, two structures $(\mathscr{S}_i, \mathbf{s}_i)$ and $(\mathscr{S}_j, \mathbf{s}_j)$ are symmetric if for some $s \in \mathfrak{R}$, $s \circ \tilde{\mathbf{s}}_i = \tilde{\mathbf{s}}_j$.

Definition A.5 (Interaction Graph)**.** *Let $\mathcal{S}$ be a set of normalised structures and let $\mathcal{T}$ be a set of triples (S_i, S_j, v) where $S_i, S_j \in \mathcal{S}$ and $v \in \mathbb{Z}^D$. Let $\mathfrak{s} : \mathcal{T} \to \mathcal{S}$, $\mathfrak{t} : \mathcal{T} \to \mathcal{S}$, and $\tau : \mathcal{T} \to \mathcal{C}$, where $\mathcal{C}$ is an arbitrary finite set. Then the $\mathcal{C}$-decorated multidigraph $\mathcal{I} = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ is the* interaction graph *of an autopoietic process dependency graph G when:*

1. *$\mathcal{S}$ satisfies conditions (1)–(4) in* Definition A.4 *for every $(\mathscr{S}, \mathbf{s}) \in \mathcal{S}$ with respect to G;*
2. *$\mathcal{T}$ gives all $(S_i, S_j, \cdot)$ that satisfy condition (3) for G;*
3. *Let $S_i, S_j \in \mathcal{S}$ be structures. If $\tau(t \in \mathcal{T}) = n$, then $\tau(t') = n$ iff there exists some fixed $s \in \mathfrak{R}$ such that $s \circ \mathbf{s}_i = \mathbf{s}'_i$, $s \circ \mathbf{s}_j = \mathbf{s}'_j$, and $s(v) = s(v')$.*

The procedure for deriving an interaction graph from an autopoietic process dependency graph is fairly involved. First, we need a way to extract forward fragments from a given process. I will represent fragments as pairs $F_i = (f_i, \vec{F}_i)$, where $f_i : V \to \mathscr{F}_i \cup \{0\}$ is a map from vertices to a set of coordinates $\mathscr{F}_i \subset \mathbb{Z}^D$ or 0, and

$$\vec{F}_i(x) := \bigcup_{j \in f_i^{\leftarrow}} \sigma^{f_i(j)} \circ \mathbf{p}_j(x). \tag{A.5}$$

Then Algorithm 1 can be used to derive a set of fragments enabled by a given process. By taking the forward fragments of every process in a graph, we can merge them all until we get a full set of process embeddings that correspond to structures (Algorithm 2). To get the edges of the interaction graph, we need an algorithm that derives $\mathcal{T}$ from a set of *normalised* process embeddings $\mathfrak{S}$ and a dependency graph G (Algorithm 3). An embedding is normalised by simply centring the codomain of f_i about $\vec{0}$ (or any other fixed reference point). Getting the source and target maps is straightforward given our representation of edges: $\mathfrak{s}(t \in \mathcal{T}) = t_i$ and $\mathfrak{t}(t \in \mathcal{T}) = t_j$, where t_i gives S_i from $(S_i, S_j, v) \in \mathcal{T}$, and similarly

Algorithm 1 LtL Fragment Extraction Algorithm

```
Require: $G = (V, E, \mathfrak{s}, \mathfrak{t})$ is a process dependency graph (Definition A.3) and $i \in V$ where $\{(i, \cdot, \cdot) \in E\}$ is non-empty.
  $E_i \leftarrow \{(i, \cdot, \cdot) \in E\}$
  $\mathfrak{F} \leftarrow \{(f_j, \mathbf{p}_j) : (i, j, (\vec{0}, \cdot, \cdot)) \in E_i\}$ s.t. $f_j(j) = \vec{0}$
  for all $r \in \mathscr{N} \setminus \{\vec{0}\}$ do
      for all $F \in \mathfrak{F}$ do
          for all $\psi_j \in \{\psi_j : (i, j, (r, \cdot, \cdot)) \in E_i \setminus, \psi_j \notin \mathrm{supp}(\mathfrak{F})\}$ do
              $\vec{F}' \leftarrow \vec{F} \cup (\sigma^{-r} \circ \mathbf{p}_j)$
              if $\vec{F}'$ is well-defined, $f^{-1}(-r)$ is not defined, and $\exists a \in \mathrm{supp}(f)$ s.t. $-r - f(a) = r_1 - r_2$ where $(j, b, (r_1, \cdot, \cdot)), (a, b, (r_2, \cdot, \cdot)) \in E$ for some $b \in V$ then
                  $f(j) \leftarrow -r$
              end if
          end for
      end for
  end for
  for all $r \in \mathscr{N} \setminus \{\vec{0}\}$ do
      for all $F \in \mathfrak{F}$ do
          for all $\psi_j \in \{\psi_j : (i, j, (r, \cdot, \cdot)) \in E' \setminus \mathrm{flatten}(\mathfrak{F})\}$ do
              $\vec{F}' \leftarrow \vec{F} \cup (\sigma^{-r} \circ \mathbf{p}_j)$
              if $\vec{F}'$ is well-defined and $f^{-1}(-r)$ is not defined then
                  $f(\psi_j) \leftarrow -r$
              end if
          end for
      end for
  end for
```

Algorithm 2 LtL Process Embedding Algorithm

```
Require: $G = (V, E, \mathfrak{s}, \mathfrak{t})$ is a process dependency graph (Definition A.3) and $\mathfrak{F}_i$ is the set
  of fragments extracted from a given $i \in V$ using Algorithm 1.
  $\mathfrak{F} \leftarrow \bigcup_{i \in V} \mathfrak{F}_i$
  $\mathfrak{S} \leftarrow \emptyset$
  $\mathfrak{D} \leftarrow \emptyset$
  for all $F_i \in \mathfrak{F}$ do
      if $F_i \notin \mathfrak{D}$ then
          $\mathfrak{S} \leftarrow \mathfrak{S} \cup \{F_i\}$
          $\mathfrak{D} \leftarrow \mathfrak{D} \cup \{F_i\}$
          $j \leftarrow 0$
          while $j < |\mathfrak{F} \backslash \mathfrak{D}|$ do
              if $\exists a \in \mathrm{supp}(f_i)$ s.t. $a \in \mathrm{supp}(f_j)$ then
                  $v \leftarrow f_i(a) - f_j(a)$
                  $\mathfrak{D} \leftarrow \mathfrak{D} \cup \{F_j\}$
                  $f_i(x) \leftarrow f_i(x) \cup f_j(x - v)$
                  $j \leftarrow 0$
              end if
          end while
      end if
  end for
```

Algorithm 3 LtL Edge Derivation Algorithm

```
Require: $G = (V, E, \mathfrak{s}, \mathfrak{t})$ is a process dependency graph (Definition A.3) and $\mathfrak{S}$ is a set of
  normalised process embeddings derived from $G$ using Algorithm 2.
  $\mathcal{T} \leftarrow \emptyset$
  for all $(F_i, F_j) \in \mathfrak{S}^2$ do
      $E' \leftarrow \{(i, j, (\vec{0}, \cdot, \cdot)) \in E : f_i(i) \neq 0, f_j(j) \neq 0\}$
      $\mathcal{T} \leftarrow \mathcal{T} \cup \{\tilde{f}_i(\imath) - \tilde{f}_j(j) : (i, j, (\vec{0}, \cdot, \cdot)) \in E'\}$
  end for
```

for t_j. The final step is to get the edge label map τ, which an be derived by considering symmetries between structures, with some additional complications to account for edges that differ only by translation (Algorithm 4).

Algorithm 4 $\mathcal{I}$ Edge Labelling

```
Require: $\mathcal{S}$ is a set of normalized structures and $\mathcal{T}$ is a set of transition between between structures.
  $\mathcal{C}' \leftarrow$ some finite set s.t. $|\mathcal{C}'| = |\mathcal{T}|$
  Let $\tau : \mathcal{T} \to \mathcal{C}'$ be a bijection
  $\texttt{used} \leftarrow \emptyset$
  for all $t, t' \in \mathcal{T}^2$ do
      $\texttt{used} \leftarrow \texttt{used} \cup \{t\}$
      if $\nexists (S'_i, \cdot, v') \in \mathcal{T})$, $t' \notin \texttt{Used}$, and $\texttt{sym?}(t, t')$ then
          $\tau(t') \leftarrow \tau(t)$
          $\texttt{used} \leftarrow \texttt{used} \cup \{t'\}$
      end if
  end for
  $\mathcal{C} \leftarrow \tau^{\to}(\mathcal{T})$
```

The function $\texttt{sym?} : \mathcal{T}^2 \to \texttt{Boolean}$ is true when there exists some $s \in \mathfrak{R}$ such that $s \circ \mathbf{s}_i \overset{\text{a.e.}}{=} \mathbf{s}'_i$ and $s \circ \sigma^v \circ \mathbf{s}_j \overset{\text{a.e.}}{=} \sigma^{v'} \circ \mathbf{s}'_j$.

In summary, then, an interaction graph $\mathcal{I}_G = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ can be derived from a process dependency graph G in the following steps:

1. Derive a set of process embeddings $\mathfrak{S}$ from G using Algorithms 1 and 2.

2. Normalise every $F \in \mathfrak{S}$.

3. Get $\mathcal{S}$ by replacing f with $\mathscr{F}$ (its codomain) for every $(\vec{F}, f) \in \mathfrak{S}$.

4. Get $\mathcal{T}$ from G and $\mathfrak{S}$ using Algorithm 3.

5. Get $\mathfrak{s}$ and $\mathfrak{t}$ from the edge representation of $\mathcal{T}$.

6. Get τ from $\mathcal{S}$ and $\mathcal{T}$ using Algorithm 4.

A.1.3. The Intrinsic Viability Constraint

State Formulation

Definition A.6 (Intrinsic Viability Constraint)**.** *Let G be an autopoietic process dependency graph by* Definition A.4 *with an interaction graph $\mathcal{I}_G = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ (*Definition A.5*). Then*

$$\mathcal{V}_G := \bigcup_{(S_i, S_j, v) \in \mathcal{T}} \left\{ \mathbf{u}\big|_{\mathcal{E}^2(\mathscr{S}_i)} \in \mathcal{U} : \mathbf{u}\big|_{\mathscr{S}_i} = \mathbf{s}_i,\ \xi[\mathbf{u}]\big|_{\mathscr{S}_j + v} = \sigma^v \circ \mathbf{s}_j \right\} \tag{A.6}$$

is the intrinsic viability constraint *of G, given some fixed LtL rule. The boundary of the constraint is defined as the set of configurations that destroy the system:*

$$\partial\mathcal{V}_G := \bigcup_{S \in \mathcal{S}} \left\{ \mathbf{u}\big|_{\mathcal{E}^2(\mathscr{S})} \in \mathcal{V}_G : \mathbf{u}\big|_{\mathscr{S}} = \mathbf{s},\ \nexists (S, S', v) \in \mathcal{T} \left(\xi[\mathbf{u}]\big|_{\mathscr{S}' + v} = \sigma^v \circ \mathbf{s}' \right) \right\} \tag{A.7}$$

Density Formulation

First, we define a function that gives constraints on Ψ given a source and target function:

$$\Gamma[\mathbf{u}_t, \mathbf{u}_{t+1}](x) = \begin{cases} [b_0, b_1] & \mathbf{u}_t(x) = 0 \wedge \mathbf{u}_{t+1} = 1 \\ [0,1] \backslash [b_0, b_1] & \mathbf{u}_t(x) = 0 \wedge \mathbf{u}_{t+1} = 0 \\ [s_0, s_1] & \mathbf{u}_t(x) = 1 \wedge \mathbf{u}_{t+1} = 1 \\ [0,1] \backslash [s_0, s_1] & \mathbf{u}_t(x) = 1 \wedge \mathbf{u}_{t+1} = 0. \end{cases} \tag{A.8}$$

We will also want a way to represent the set of density functions that satisfy a given constraint:

$$\mathcal{D}[\Gamma](\mathscr{A}) := \left\{ \Psi\big|_{\mathscr{A}} : \Psi(x) \in \Gamma(x)\ \forall\, x \in \mathscr{A} \right\}. \tag{A.9}$$

Finally, we need the multidimensional discrete-domain Fourier transform (DTFT) and its inverse:

$$\mathcal{F}[f](x) := \sum_{y \in \mathbb{Z}^D} f(y) e^{-i\langle x \cdot y \rangle} \tag{A.10}$$

$$\mathcal{F}^{-1}[f](x) := \left(\frac{1}{2\pi}\right)^D \int\limits_{[-\pi,\pi]^D} f(y) e^{i\langle x \cdot y \rangle}\, \lambda[dy] \tag{A.11}$$

where $\langle x \cdot y \rangle$ denotes the inner product $\sum_{j=1}^{n} x_j y_j$.

Conjecture A.1 (Density Formulation of Viability)**.** Let G be an autopoietic process dependency graph by Definition A.4 with an interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$. Let

$$\begin{aligned} \mathcal{D}^{ij} &:= \mathcal{D}[\Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j]](\mathscr{S}_i \cap (\mathscr{S}_j + v)) \\ U[f] &:= \mathcal{F}^{-1}\left[\frac{\mathcal{F}[f]}{\mathcal{F}[\kappa]}\right]. \end{aligned}$$

Then, given a fixed LtL rule,

$$\mathcal{V}'_G := \bigcup_{(S_i, S_j, v) \in \mathcal{T}} \bigcup_{\Psi \in \mathcal{D}^{ij}} \left\{ \mathbf{u}\big|_{\mathcal{E}(\mathscr{S}_j + v)} = U[\Psi'] : \Psi' \in \mathcal{D}[\Gamma[U[\Psi], \sigma^v \circ \mathbf{s}_j]](\mathscr{S}_j + v) \right\} \tag{A.12}$$

is equivalent to the viability constraint defined in Definition A.6 when each $\mathbf{u} \in \mathcal{V}'$ is taken to imply a set of functions over $\mathcal{E}^2(\mathscr{S}_i)$ equal over the domain of $\mathbf{u}$.

The titular point of this conjecture is U, which is a bijection from density functions over some set $\mathscr{A}$ to universe state function over $\mathcal{E}(\mathscr{A})$. Such a bijection is guaranteed to exist by the Convolution Theorem (Oppenheim & Schafer, 2010, p. 60) for the DTFT which states that, for any two *absolutely summable* discrete signals u and v (i.e., functions with $\sum_{\mathbb{Z}^D} |u| < \infty$; Oppenheim and Schafer, 2010, pp. 49–50),

$$\mathcal{F}[u * v] = \mathcal{F}[u]\mathcal{F}[v].$$

Applying this to the our case, both κ and any $\mathbf{u} \in \mathcal{U}$ over a finite domain are absolutely summable: κ is an indicator function of a finite set, and we are only considering universe state functions restricted to finite sets (we can set $\mathbf{u}(x) = 0$ for all x not in the set). Thus, we get

$$\mathcal{F}[\Psi] = \mathcal{F}[\kappa * \mathbf{u}] = \mathcal{F}[\kappa]\mathcal{F}[\mathbf{u}].$$

Some basic algebra and an inverse Fourier transform gets us

$$\mathbf{u} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\Psi]}{\mathcal{F}[\kappa]}\right]. \tag{A.13}$$

This means that every Ψ over a set $\mathscr{A}$ implies a unique function $\mathbf{u}$ over $\mathcal{E}(\mathscr{A})$.

The density formulation allows us to define two metrics of density functions with respect to a process dependency graph G.

Definition A.7 (Transition Metric). *The* transition metric*, assigning a value to how well a density function Ψ satisfies the density bounds implied by the structural transition from $\mathbf{s}_i$ to $\sigma^v \circ \mathbf{s}_j$, is defined:*

$$d_t[\Psi](x) := \inf\left\{|\Psi(x) - y| : y \in \Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j](x)\right\}. \tag{A.14}$$

Note that Γ may in part depends on Ψ, by virtue of Ψ being used to extend $\mathbf{s}_i$ over all of $\mathscr{S}_j$. One could thus use d_t over $\mathscr{S}_i \cap \mathscr{S}_j$ with a fixed Γ, and then extend the metric over $\mathscr{S}_j$.

Definition A.8 (Viability Metric). *The* viability metric*, assigning a value to how close a density function is to corresponding to any viable structural transition from S_i, is defined:*

$$d_{\mathcal{V}}[S_i, \Psi](x) := \min\{d_t[\Psi](x) : t = (S_i, \cdot, \cdot) \in \mathcal{T}\} \tag{A.15}$$

Much like the previous metric, Γ can only be fixed over some intersection with $\mathscr{S}_i$. And again, a choice of Ψ can extend the metric over a larger region. Note that $d_{\mathcal{V}} = 0$ does not imply $d_t = 0$, since the parts of $d_{\mathcal{V}}$ that are satisfied may correspond to different Γ. Thus, $d_{\mathcal{V}} = 0$ does not mean that a perturbation is actually viable, but that it satisfies the bounds of at least one transition at every point.

A.1.4. Rule Derivations

Section A.1.3 assumed a fixed LtL rule $(\rho, b_0, b_1, s_0, s_1)$. However, it is possible to derive a set of rules under which an organisation can be realised, assuming it is reversible. The following procedure is non-exhaustive: for certain organisations, there may be rules that support it but are do not satisfy the constraints derived from the procedure.

The set of rules can be implicitly specified by an 8-tuple $(b_0^L, b_0^U, b_1^L, b_1^U, s_0^L, s_1^U)$, where

$$b_0^L < b_0 \leq b_0^U \tag{A.16}$$

$$b_1^L \leq b_1 < b_1^U \tag{A.17}$$

$$s_0^L < s_0 \leq s_0^U \tag{A.18}$$

$$s_1^L \leq s_1 < s_1^U \tag{A.19}$$

$$0 < s_0 \leq b_0 \leq b_1 \leq s_1 \leq 1. \tag{A.20}$$

Note that this last constraint is a modification of the general LtL rule constraint: here, $b_0 > 0$ must be true in order for vacuums to be possible.

To get values for these parameters, we first partition V into four sets:

$$\mathcal{X}^b := \left\{ \psi_i : \mathbf{p}_i(\vec{0}) = 0 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 1) \right\} \tag{A.21}$$

$$\mathcal{X}^{\neg b} := \left\{ \psi_i : \mathbf{p}_i(\vec{0}) = 0 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 0) \right\} \tag{A.22}$$

$$\mathcal{X}^s := \left\{ \psi_i : \mathbf{p}_i(\vec{0}) = 1 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 1) \right\} \tag{A.23}$$

$$\mathcal{X}^{\neg s} := \left\{ \psi_i : \mathbf{p}_i(\vec{0}) = 1 \wedge \exists (i, \cdot, k) \in E(k_{c_1} = 0) \right\}. \tag{A.24}$$

I am assuming here that, for every $i \in V$, k_{c_1} is constant over all $(i, \cdot, k) \in E$. This is implied by the interaction graph property discussed above.

Since I am only considering vacuum perturbations, we can define Ψ over a partial process by assuming that for every ψ, $\mathbf{p}(x) = 0$ for all $x \in \overline{\phi}$. Then the parameters are defined

$$b_0^U := \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} \tag{A.25}$$

$$b_0^L := \max\left\{\Psi(\vec{0}) : \Psi(\vec{0}) < b_0^U,\ \psi \in \mathcal{X}^{\neg b}\right\} \tag{A.26}$$

$$b_1^L := \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^b\right\} \tag{A.27}$$

$$b_1^U := \min\left\{\Psi(\vec{0}) : \Psi(\vec{0}) > b_1^L,\ \psi \in \mathcal{X}^{\neg b}\right\} \tag{A.28}$$

$$s_0^U := \min\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} \tag{A.29}$$

$$s_0^L := \max\left\{\Psi(\vec{0}) : \Psi(\vec{0}) < s_0^U,\ \psi \in \mathcal{X}^{\neg s}\right\} \tag{A.30}$$

$$s_1^L := \max\left\{\Psi(\vec{0}) : \psi \in \mathcal{X}^s\right\} \tag{A.31}$$

$$s_1^U := \min\left\{\Psi(\vec{0}) : \Psi(\vec{0}) > s_1^L,\ \psi \in \mathcal{X}^{\neg s}\right\}, \tag{A.32}$$

These definitions need some qualifications. If any of the values do not exist (i.e., when one of the sets $\mathcal{X}$ is empty), they can be ignored in the above inequalities. If such a parameter is needed to compute another one, more care is needed. When $\mathcal{X}^b$ is empty, ignore the b_0^U constraint in the definition of b_0^L; b_1^U can be ignored entirely. Otherwise, one can safely ignore the constraint in a parameter definition when that constraint is undefined.

The logic of this is grounded in the following. First, $\mathcal{X}^{\neg b}$ must always contain a partial process that produces a 0-component due to being sufficiently *below* the b_0 boundary. This is an assumption that is not *always* necessary, but making it allows the calculations to be made (it also corresponds to a condition that *is* necessary in Section A.2.4). Second, b_1^U only exists when there are process that remain off due to sufficiently *high* Ψ. I assume that such process can only exist when $\mathcal{X}^b$ is non-empty: that there must be an intermediate Ψ that results in a production process between a Ψ above and below $[b_0, b_1]$. Again, this is not strictly necessary in LtL, but corresponds to a necessary condition in RealLife. Finally, I assume that $\mathcal{X}^s$ is non-empty.

This procedure should only derive rules that support the realisation of an organisation as some periodic LtL pattern. When any of the sets $\mathcal{X}$ do not exist, the procedure may not be exhaustive. However, in the more general case where they do exist, I conjecture that this procedure will derive all and only those LtL rules that support the asymptotic realisation of an organisation in a vacuum universe. Stated more formally:

Conjecture A.2 (Sufficiency of the LtL Rule Derivation)**.** Let $G = (V, E, \mathfrak{s}, \mathfrak{t})$ be a reversible autopoietic process dependency graph.

I. All rules that satisfy the constraints derived by the procedure just described are such that, for every $(S_i, S_j, v) \in \mathcal{T}$:

$$\xi[\mathbf{s}_i] = \sigma^v \circ \mathbf{s}_j \tag{A.33}$$

where $\mathbf{s}(x) = 0$ for all $x \notin \mathcal{S}$.

II. When $\mathcal{X}^b$, $\mathcal{X}^{\neg b}$, $\mathcal{X}^s$, and $\mathcal{X}^{\neg s}$ all exist, the set of rules that satisfy the derived constraints contains *every* rule such that, for every $(S_i, S_j, v) \in \mathcal{T}$, Equation A.33 is true.

A.2. RealLife

Let $\mathcal{M} := \{0, 1\}$ and let λ be the D-dimensional Lebesgue measure on $\mathbb{R}^D$. A *universe state function* is a Borel-measurable function $\mathbf{u} : \mathbb{R}^D \to \mathcal{M}$. We denote the space of all universe states $\mathcal{U} := \mathcal{M}^{\mathbb{R}^D} \subset \mathcal{L}^\infty(\mathbb{R}^D, \lambda)$. Let the *neighbourhood* be a set $\mathcal{N} := \{x : \|x\|_p \leq \rho\}$ centred at the origin, where $\rho \in \mathbb{R}^+$ is a finite neighbourhood radius.

Let $\kappa(x) := \lambda[\mathcal{N}]^{-1} \mathbb{1}_{\mathcal{N}}(x)$ be a convolution kernel. The *density function* $\Psi : \mathbb{R}^D \to [0, 1]$ is defined as a convolution, giving the average value about a point x:

$$\Psi(x) := (\kappa * \mathbf{u})(x) = \lambda[\mathcal{N}]^{-1} \int\limits_{y \in \mathcal{N}} \mathbf{u}(x - y)\, \lambda[dy] \tag{A.34}$$

A *D-dimensional RealLife Euclidean automaton* (Pivato, 2007) is a functional $\xi : \mathcal{U} \to \mathcal{U}$, defined

$$\xi[\mathbf{u}](x) := \mathbf{u}(x) \mathbb{1}_{[s_0, s_1]}(\Psi(x)) + (1 - \mathbf{u}(x))\, \mathbb{1}_{[b_0, b_1]}(\Psi(x)), \tag{A.35}$$

where $0 \leq s_0 \leq b_0 < b_1 \leq s_1 \leq 1$. Thus a RealLife rule can be fully specified by a 5-tuple $(\rho, s_0, b_0, b_1, s_1)$ and a choice of p-norm.

It will be useful to define a few additional sets and functions. Let $\sigma^v[\mathbf{u}](x) := \mathbf{u}(x - v)$ be a functional that translates state functions for some $v \in \mathbb{R}^D$. Let $\mathcal{E}(\mathscr{A}) := \bigcup_{x \in \partial \mathscr{A}} \mathscr{N} + x$ give the set $\mathscr{A} \subset \mathbb{R}^D$ and its 1-environment. We get the n-environment by recursion: $\mathcal{E}^n$. With a slight abuse of notation, we will similarly use $\mathcal{E}^n[\mathbf{s}]$ to denote the set of functions over $\mathcal{E}^n(\mathscr{A})$ that are equal over $\mathscr{A}$. Finally, let $\mathscr{N}_{op} := \left\{x : \|x\|_p < \rho, x \in \mathbb{R}^D\right\}$ be the open neighbourhood about the origin.

A.2.1. Autopoiesis

Definition A.9 (Processes). *A* process *in RealLife is a tuple* $(\phi, \mathbf{p})$ *where* $\phi \subseteq \mathscr{N} + v$ *for some finite* $v \in \mathbb{R}^2$, *and* $\mathbf{p}(x)$ *is almost well-defined over* ϕ. *Two processes,* $(\phi_i, \mathbf{p}_i)$ *and* $(\phi_j, \mathbf{p}_j)$, *are* chemically equivalent *if, when translated to the origin (*$\phi \mapsto \phi - v$*):*

1. ***(λ-equivalence of space)*** $\lambda[\phi_i \backslash \phi_j] = \lambda[\phi_j \backslash \phi_i]$,

2. ***(λ-equivalence of state)*** $\text{ess supp}(\mathbf{p}_i) \cap \phi_i = \text{ess supp}(\mathbf{p}_j) \cap \phi_j$.

3. ***(Equivalence of type)*** $\mathbf{p}_i(\vec{0}) = \mathbf{p}_j(\vec{0})$.

Definition A.10 (Structures). *Let* $\mathscr{S} \subset \mathbb{R}^D$ *be a finite, closed, and connected set and let* $\mathbf{s} \in \mathcal{U}$ *be a function well-defined for almost all* $x \in \mathscr{S}$. *Then* $(\mathscr{S}, \mathbf{s})$ *is a valid* structure *if* $\lambda[\text{supp}(\mathbf{s})] > 0$, $\mathscr{S} = \bigcup_{x \in \text{ess supp}(\mathbf{s})} \mathscr{N} + x$, *and* $\mathbf{s}(x) = 0$ *for almost all* $x \in \mathscr{S} \cap \bigcup_{y \in \partial \mathscr{S}} \mathscr{N}_{op} + y$. *Two structures* $(\mathscr{S}_i, \mathbf{s}_i)$ *and* $(\mathscr{S}_j, \mathbf{s}_j)$ *are equivalent if they are almost translationally symmetric, that is, if there exists some* $v \in \mathbb{R}^D$ *such that* $\mathscr{S}_i + v \overset{\text{a.e.}}{=} \mathscr{S}_j$ *and* $\sigma^v \circ \mathbf{s}_i \overset{\text{a.e.}}{=} \mathbf{s}_j$.

With these, we can now define an autopoietic system as a class of process dependency networks by specifying the conditions that such a network must satisfy to be autopoietic. The following definition combines and adapts various definitions from the theory of graph limits, including graphons on probability spaces (Lovász, 2012, p. 217), K-graphons (Lovász, 2012, p. 322), and digraphons (Boeckner, 2013, pp. 14–15, 32–36).

Definition A.11 (Process Dependency Graphon). *Let* $J = (\Omega, \mathcal{A}, \pi)$ *be a probability space and* $K := \mathscr{N} \times \mathscr{M}^2$ *a compact Hausdorff space. Let* $\omega : J \times J \to \mathbb{P}(K_0)$ *be a function measurable with respect to the completion of the sigma-algebra* $\mathcal{A} \times \mathcal{A}$*, where* $\mathbb{P}(\mathscr{A})$ *is the space of probability measures over the set* $\mathscr{A}$*. Let* k_{c_0} *give the* c_0 *(source) and* k_{c_1} *give the* c_1 *(target)* $\mathscr{M}$ *component for all* $k \in K$*; let* k_r *give the* $\mathscr{N}$ *component. Then* ω *is a valid process dependency graphon if, for every* $x \in \Omega$*:*

1. *The codomain of* ω *is the codomain of some bijection* $g : K_0 \to \mathbb{P}(K_0)$*;*

2. k_{c_1} *is constant over all* $y \in \{y : (g^{-1} \circ \omega)_r(y,x) = k_r\}$ *for every fixed* k_r *where* $g^{-1} \circ \omega(x,y) \in K$*;*

3. k_{c_0} *is constant over all* y *where* $g^{-1} \circ \omega(x,y) \in K$*; and*

4. *For every* $S \subseteq \Omega$ *with* $0 < \lambda[S] < \lambda[\Omega]$*,*

$$\int_{S\times(\Omega\backslash S)} \Big(\mathbb{1}_K\big[(g^{-1} \circ \omega)(y_1,y_2)\big] + \mathbb{1}_K\big[(g^{-1} \circ \omega)(y_2,y_1)\big] \Big)\, dy_1 dy_2 > 0. \tag{A.36}$$

Two process dependency graphons are equivalent if there exists a measure-preserving bijection $f : \Omega_i \to \Omega_j$ *such that for all* $x,y \in \Omega_i$*,* $\omega(x,y) \overset{\text{a.e.}}{=} \omega(f(x),f(y))$*.*

ω *is* closed *if, for every* $x \in \Omega$*, there exists both a* $y \in \Omega$ *such that* $g^{-1} \circ \omega(x,y) \in K$ *and a* $y' \in \Omega$ *such that* $g^{-1} \circ \omega(y,x) \in K$*.*

ω *is* reversible *if, for every* $x \in \Omega$ *and* $r \in \mathscr{N}$ *where there exists some* y *such that* $g^{-1} \circ \omega_r(y,x) = r$*, there is only one such* y*.*

ω *is* degenerate *if there exists some* $x \in \Omega$ *such that (i) there is no* y *where* $g^{-1} \circ \omega_r(y,x) = \vec{0}$*, and (ii) there is no* y' *where* $g^{-1} \circ \omega(y,x) \in K$*.*

Thus, process dependency kernels are directed K_0-graphons. Equation A.36 is an adaptation of the connectivity of symmetric kernels (Lovász, 2012, p. 122) to define *weakly connected* process dependency networks.

Here, Ω represents the set of processes in the network and K represents the space of enabling relations as a triple (r, c_0, c_1) of the center-component c_0 of the source process, the component produced c_1, and the position r of the produced component in the target process; $0 \in K_0$ is the lack of such a relation. ω is a function that assigns such relations to pairs of processes. The use of probability measures is needed in Section A.3, but it will in general be more useful to deal with $W := g^{-1} \circ \omega$.

The reader will notice the use of equivalence up to measure. This is because the RealLife universe has subsets of zero measure, but Ψ is only effected by measurable sets, and so any subset of zero measure cannot influence dynamics outside that set. In other words, any modification to a structure that is equivalent up to measure will have effectively the same dynamics. I therefore do not consider these structural variations worth representing in the graphon: only a single representative from an equivalence class of structures needs to be derived from the graphon in order for it to specify all possible structural realisations of itself.

Processes are implicit in this definition, but can still be extracted. Given an $x \in \Omega$, $\phi_x = \{W_r(y,x) : y \in \Omega\}$. To get $\mathbf{p}_x$, we first need to define an inverse: $\overleftarrow{W}^z(x) := \{y : W_r(y,x) = z\}$. Then $\mathbf{p}_x(z) = W_{c_1}(\overleftarrow{W}^z(x), x) \cup W_{c_0}(x, \cdot)$, which is guaranteed to be consistently defined by condition (2) of Definition A.11.

Finally, we can specify when a process dependency graphon is autopoietic.

Definition A.12 (Autopoiesis). *A process dependency graphon ω is autopoietic if there exists a partition $\{O_i\}$ of Ω such that every O_i satisfies the following conditions:*

1. *There exists a bijection $f_i : O_i \to \mathscr{S}_i$ such that $(cl_{\mathbb{R}^2}\mathscr{S}_i, \mathbf{p}_{f_i^{-1}}(\langle 0,0\rangle))$ is a valid structure.*

2. *$(\mathscr{S}_i, \mathbf{s}_i)$ can be expressed as the union of translated processes in O_i.*

3. *There exists O_j such that for some $\mathbf{u} \in \mathcal{E}[\mathbf{s}_i]$ and a RealLife rule, $\xi[\mathbf{u}]|_{\mathscr{S}_i \cap (\mathscr{S}_j + v)} = \sigma^v \circ \mathbf{s}_j$, where $v = f_i(x_i) - f_j(x_j)$ such that $x_i \in O_i$, $x_j \in O_j$, and $W_r(x_i, x_j) = \vec{0}$. Let $\mathscr{D}_{ij} = \mathscr{S}_i \cap (\mathscr{S}_j + v)$. Then for every $f_i(x_i) \in \mathscr{D}_{ij}$, $W_r(x_i, x_j) = f_j(x_j) - f_i(x_i) - v$ for all x_j such that $f_j(x_j) + v \in (\mathscr{N} + f_i(x_i)) \cap (\mathscr{S}_j + v)$.*

4. *There exists O_j such that (3) is satisfied where O_i and O_j are swapped.*

The network as a whole must also be such that all $\langle x_i, x_j \rangle \in \operatorname{supp}(W)$ *satisfy* $f_i(x_i) \in \mathscr{D}_{ij}$ *and* $W_r(x_i, x_j) = f_j(x_j) - f_i(x_i) - v_{ij}$ *for some* O_i, O_j *valid according to condition (3).*

According to this definition, then, an autopoietic process dependency graphon implies a closed network of structures and every enabling relation participates in some transition between structures. Note that the graphon itself need not be *completely* closed, but that every component that constitutes the individual at some point is specified by the network. It may be possible to prune from the network those processes that do not enable any other process and still recover complete structures where they are implicit (as can be done quite easily in the discrete case). However, this would require a more complicated statement that combines conditions (1) and (2), or else replace conditions (1)-(4) with "must derive a closed network of structure transitions" using the procedure outlined in Section A.2.2. As such, I will use the simpler definition of autopoiesis.

A.2.2. The Cognitive Domain

Given that I am considering only autopoietic systems with finitely many structures, the proper object to describe the cognitive domain is an *interaction graph*: a digraph with structures as vertices and transition vectors as edges.

First, we must define the symmetries of a structure as a dihedral group of degree 4 and order 8, D_8. This group can be generated by two functionals representing rotations, $\mathbf{r}$, and flips, $\mathbf{f}$. To define these, we need to normalise structure representations. Let $(\tilde{\mathscr{S}}, \tilde{\mathbf{s}})$ denote the normalised representation of a structure such that $\tilde{\mathscr{S}} := \mathscr{S} - v$ and $\tilde{\mathbf{s}} := \sigma^{-v} \circ \mathbf{s}$, where

$$v = \Bigg\langle \inf\{x : \exists y \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\} - \frac{\sup\{x : \exists y \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\} - \inf\{x : \exists y \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\}}{2},$$
$$\inf\{y : \exists x \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\} - \frac{\sup\{y : \exists x \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\} - \inf\{y : \exists x \text{ s.t. } \langle x,y \rangle \in \mathscr{S}\}}{2} \Bigg\rangle.$$

Then

$$\mathbf{f}[\tilde{\mathbf{s}}]\,(x) := \mathbf{s}(\langle -x_0, x_1\rangle) \tag{A.37}$$

$$\mathbf{r}[\tilde{\mathbf{s}}]\,(x) := \mathbf{s}\left(\left\langle \|x\|_2 \cos\left(\frac{\pi}{2} + \tan^{-1}\frac{x_1}{x_0}\right), \|x\|_2 \sin\left(\frac{\pi}{2} + \tan^{-1}\frac{x_1}{x_0}\right)\right\rangle\right) \tag{A.38}$$

generates the group $\mathfrak{R} = \langle\{\mathbf{f}, \mathbf{s}\}\rangle$, which is isomorphic to D_8. Thus, two structures $(\mathscr{S}_i, \mathbf{s}_i)$ and $(\mathscr{S}_j, \mathbf{s}_j)$ are symmetric if for some $s \in \mathfrak{R}$, $s \circ \tilde{\mathbf{s}}_i = \tilde{\mathbf{s}}_j$.

Definition A.13 (Interaction Graph)**.** *Let $\mathcal{S}$ be a set of structures and let $\mathcal{T}$ be a set of triples (S_i, S_j, v) where $S_i, S_j \in \mathcal{S}$ and $v \in \mathbb{R}^2$. Let $\mathfrak{s} : \mathcal{T} \to \mathcal{S}$, $\mathfrak{b} : \mathcal{T} \to \mathcal{S}$, and $\tau : \mathcal{T} \to \mathcal{C}$, where $\mathcal{C}$ is an arbitrary finite set. Then the $\mathcal{C}$-decorated multidigraph $\mathcal{I} = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ is an* interaction graph *of an autopoietic process dependency graphon ω when:*

1. *$\mathcal{S}$ satisfies conditions (1)–(4) in* Definition A.12 *for every $(\mathscr{S}, \mathbf{s}) \in \mathcal{S}$ with respect to ω;*
2. *$\mathcal{T}$ gives all $(S_i, S_j, \cdot)$ that satisfy condition (3) for ω;*
3. *Let $S_i, S_j \in \mathcal{S}$ be structures. If $\tau(t \in \mathcal{T}) = n$, then $\tau(t') = n$ iff there exists some fixed $s \in \mathfrak{R}$ such that $s \circ \tilde{\mathbf{s}}_i = \tilde{\mathbf{s}}'_i$, $s \circ \tilde{\mathbf{s}}_j = \tilde{\mathbf{s}}'_j$, and $s(v) = s(v')$.*

Given an autopoietic process dependency graphon ω, an interaction graph $\mathcal{I}_\omega$ can be derived using Algorithm 5. However, we will need to define a few things first.

Let $\overrightarrow{W}^z(x) := \{y : W_r(x, y) = z\}$. Then we define a *forward fragment* of a process x as a pair $(\overrightarrow{\mathscr{F}_x}, \overrightarrow{\mathbf{F}}_x)$, where:

$$\overrightarrow{\mathscr{F}_x} := \bigcup_{z \in \mathscr{N}} \phi_{\overrightarrow{W}^z(x)} - z \tag{A.39}$$

$$\overrightarrow{\mathbf{F}}_x := \bigcup_{z \in \mathscr{N}} \sigma^{-z} \circ \mathbf{p}_{\overrightarrow{W}^z(x)}. \tag{A.40}$$

Note that this function is not well-defined, since it is possible to have some $y_1, y_2 \in \Omega$ such that $W_{c_1}(x, y_1) \neq W_{c_1}(x, y_2)$ while $W_r(x, y_1) = W_r(x, y_2)$. We will therefore want a way to get multiple functions from $\vec{F}$ so that each is well-defined. It is unclear to me how to do this in general, or whether such is possible without assuming a topology on Ω. However, we can specify a partition on ω (and, thus, an induced topology) that allows us to separate these fragment functions with finite indices. Such a partition, $\{\mathcal{O}_i\}$, must be such that it is finite and for every $x \in \Omega$:

1. If $W_r(x,y_1) = W_r(x,y_2)$ and $y_1 \neq y_2, y_1 \in O_i$ and $y_2 \in O_j$, where $i \neq j$,
2. For any pair $y_1,y_2 \in O_i$ such that $W_r(x,y_1) = z_1$, $W_r(x,y_2) = z_2$, and $y_1 \neq y_2$, $\mathbf{p}_{y_1}(a - z_1) = \mathbf{p}_{y_2}(a - z_2)$ for all $a \in (\phi_{y_1} + z_1) \cap (\phi_{y_2} + z_2)$,
3. $|\{i : W(x,y) \neq 0,\, y \in O_i\}| = \max\{|W_r(x,y)| : y \in \Omega\}$,

and similarly for $W_r(\cdot,x)$. We will say that if a graphon ω has such a partition, Ω has the *fragment partition topology*. It turns out that this partition separates processes by what structure they belong to, since the relation of belonging to the same fragment is transitive. However, no topology is assumed *within* a given partition, so work still needs to be done to recover the structural realisations of ω. In any case, I have tried to minimise dependence on the partition by not assuming that two fragment belong to the same set until such can be shown independent of the partition.

We can thus redefine forward fragments:

$$\overrightarrow{\mathscr{F}}_{x,i} := \bigcup_{z\in\mathscr{N}} \phi_{\overrightarrow{W}^z(x)\cap O_i} - z \tag{A.41}$$

$$\overrightarrow{\mathbf{F}}_{x,i} := \bigcup_{z\in\mathscr{N}} \sigma^{-z} \circ \mathbf{p}_{\overrightarrow{W}^z(x)\cap O_i}. \tag{A.42}$$

When a partition is not specified, we get a set of functions $\{\mathbf{F}_i\}$ such that for every i, there exists $y \in O_i$ where $W(x,y) \neq 0$.

We will need to keep track of a well chosen set of points in each fragment. Let $P_i \in \mathfrak{P}$ be a set of pairs $(x' \in \overrightarrow{\mathscr{F}_i}, x'' \in \Omega)$ for the fragment $\vec{F}_i$. Algorithm 5 will show how to get this set, but it will require another bit of machinery: Let $\mathcal{K} \subset \partial\mathscr{N}$ be a set of vectors such that $\mathcal{K}$ is finite and

$$\left(\bigcup_{r\in\mathcal{K}} \mathscr{N} + r\right) \setminus \overrightarrow{\mathscr{F}} \neq \emptyset.$$

We will also need to use a metric for how well a fragment matches with a structure. Let $\vec{F}_x$ be a fragment and S_i be a structure (a completed fragment). Then define the *fragment-structure match* between $\vec{F}_x$ and S_i as:

$$\epsilon_v\left[\vec{F}_x, S_i\right] = \vec{\mathbf{F}}_x\Big|_{\mathscr{S}_i\cap(\overrightarrow{\mathscr{F}}_x+v)} - \mathbf{s}\Big|_{\mathscr{S}_i\cap(\overrightarrow{\mathscr{F}}_x+v)} = 0, \tag{A.43}$$

where $v \in \mathbb{R}^D$ is some displacement vector for the fragment into the structure. To determine whether $\vec{F}_x$ and S_i can be merged, we will need to determine whether they share a process. Let $\mathcal{M}(\vec{F}_i, S_j)$ be a boolean function that indicates whether there is a shared process. That is, it will displace all $x' \in \overrightarrow{\mathscr{F}_i}$ by v, where $\epsilon_v[\vec{F}_i, S_j] = 0$, and similarly for all $(x_i', \cdot) \in P_i$ and $(x_j', \cdot) \in P_j$ (if there is no v such that $\epsilon_v = 0$, $\mathcal{M}$ is false). Then, if there exists some $(x', x'') \in P_i$ and $(y', y'') \in P_j$ such that there is some $z \in \Omega$ with $W_r(z, x'') - W_r(z, y'') = y' - x'$, $\mathcal{M}$ is true.

Finally, we will need to define a terminating condition. Let $\mathfrak{F}$ denote a set of fragments and let $\mathfrak{F}^* \subseteq \mathfrak{F}$ denote the subset of complete fragments (i.e., structures by Definition A.10). Then, if for any $\vec{F}_i \in \mathfrak{F}^*$ there is some $\vec{F}_x$, with $(\cdot, x) \in P_i$ such that there does not exist a $\vec{F}_j \in \mathfrak{F}^*$ where $\mathcal{M}(\vec{F}_i, \vec{F}_j)$ is true, then the algorithm continues. Let $T_\omega(\mathfrak{F}, \mathfrak{P})$ denote the boolean value of the terminating condition (where false means the algorithm continues).

With simple modifications to Algorithm 5, an algorithm for non-reversible organisations can be defined, assuming those organisations have the fragment partition topology. Note that there are obvious inefficiencies in this algorithm, but it will complete in finite time so long as the derivation is possible.

Using Algorithm 5, we can derive an interaction graph from an autopoietic process dependency graphon ω. All that is left to do is define $\mathcal{I}$ in terms of ω, $\mathfrak{F}$, and $\mathfrak{P}$. First and most simply, $\mathcal{S} = \left\{(\tilde{\overrightarrow{\mathscr{F}}}, \tilde{\vec{\mathbf{F}}}) : (\overrightarrow{\mathscr{F}}, \vec{\mathbf{F}}) \in \mathfrak{F}\right\}$; coordinates in $\mathfrak{P}$ should also be shifted accordingly. Then, for each $S_i \in \mathcal{S}$, extract the forward fragments of every x_i'' where $(x_i', x_i'') \in P_i$. For each such fragment, if $\epsilon_v = 0$ for some $S_j \in \mathcal{S}$, add $(S_i, S_j, x_j' - x_i')$ to $\mathcal{T}$ (without duplicates) where $W_r(x_i'', x_j'') = \vec{0}$ and $(x_j', x_j'') \in P_j$. To get the source and target maps, let

Algorithm 5 RealLife Structure Derivation for Reversible Organisations

```
Require: $\omega$ is autopoietic by Definition A.12 and reversible.
  $x_0 \leftarrow$ some $x_0 \in \Omega$ s.t. $W(x,y) \neq 0$ for all $y \in X$ where $\lambda[X] > 0$.
  $\vec{F}_i \leftarrow \vec{F}_{x_0}$
  $\mathfrak{F} \leftarrow \{\vec{F}_i\}$
  $P_i \leftarrow \{(\vec{0}, \overrightarrow{W}^{\vec{0}}(x_0))\}$
  $\mathfrak{P} \leftarrow \{P_i\}$
  while $T_\omega(\mathfrak{F}, \mathfrak{P})$ is false do
    if $\vec{F}_i$ satisfies Definition A.10 then
      $x \leftarrow x$ s.t. $(\vec{0}, x) \in P_i$ and $\neg\mathcal{M}(\vec{F}_x, S)$ for all $S \in \mathfrak{F}$
      $\mathfrak{F} \leftarrow \mathfrak{F} \cup \vec{F}_x$
      $\mathfrak{P} \leftarrow \mathfrak{P} \cup P_x$
      $\vec{F}_i \leftarrow \vec{F}_x$
      $P_i \leftarrow P_x$
    else
      $X \leftarrow \bigcup_{(x',x'') \in P_i} \{(x_r, x'+r) \in \Omega : W(x_r, x'') = (\cdot, \cdot, -r), r \in \mathcal{K}\}$
      $P_i \leftarrow P_i \cup \bigcup_{(r_x,x_r) \in X} \left(r_x, \overrightarrow{W}^{\vec{0}}(x_r)\right)$
      $\vec{F}_i \leftarrow \vec{F}_i \cup \bigcup_{(r_x,x_r) \in X} \left(\overrightarrow{\mathscr{F}}_{x_r} + r_x, \sigma^{r_x} \circ \overrightarrow{\mathbf{F}}_{x_r}\right)$
    end if
  end while
```

$\mathfrak{s}(t \in \mathcal{T}) = t_i$ and $\mathfrak{t}(t \in \mathcal{T}) = t_j$, where t_i gives S_i from $(S_i, S_j, v) \in \mathcal{T}$, and similarly for t_j.

Algorithm 4 describes the procedure for deriving an appropriate $\tau : \mathcal{T} \rightarrow \mathcal{C}$. The function `sym?` : $\mathcal{T}^2 \rightarrow$ `Boolean` is true when there exists some $s \in \mathfrak{R}$ such that $s \circ \tilde{\mathbf{s}}_i \overset{\text{a.e.}}{=} \tilde{\mathbf{s}}'_i$ and $s \circ \sigma^v \circ \mathbf{s}_j \overset{\text{a.e.}}{=} \sigma^{v'} \circ \mathbf{s}'_j$. Finally, we get the $\mathcal{C}$-decorated multidigraph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, a, b, \tau)$.

I end this section with two conjectures claiming that, on the basis of Algorithm 5 and the procedures just described, one can *always* derive an interaction graph from an autopoietic graphon ω whenever fragments can be separated.

Conjecture A.3 (Constitutive-Interactive Relation)**.** Let ω be an autopoietic process dependency graphon by Definition A.12. Then an interaction graph $\mathcal{I}_\omega$ (Definition A.13) can be derived from it when, for all $x \in \Omega$, there is a derivable partition $\{O_i\}$ of $\bigcup_{z \in \mathscr{N}} \overrightarrow{W}^z(x)$ such that $|\{O_i\}| = \max\{|\overrightarrow{W}^z(x)| : z \in \mathscr{N}\}$ and

$$\vec{F}_x(y) = \bigcup_{z \in O_i} \mathbf{p}_z(y)$$

is well-defined for every O_i.

Conjecture A.4 (Sufficiency of the Fragment Partition Topology)**.** Let ω be an autopoietic process dependency kernel by Definition A.12. Then an interaction graph $\mathcal{I}_\omega$ (Definition A.13) can always be derived using Algorithm 5 whenever ω has the fragment partition topology. That is, an autopoietic process dependency graphon with the fragment partition topology satisfies the conditions of Conjecture A.3.

A.2.3. The Intrinsic Viability Constraint

State Formulation

Definition A.14 (Intrinsic Viability Constraint)**.** *Let ω be an autopoietic process dependency kernel by* Definition A.12 *with an interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ (*Definition A.13; Conjecture A.3*). Then*

$$\mathcal{V}_\omega := \bigcup_{(S_i, S_j, v) \in \mathcal{T}} \left\{ \mathbf{u}\big|_{\mathcal{E}^2(\mathscr{S}_i)} \in \mathcal{U} : \mathbf{u}\big|_{\mathscr{S}_i} \overset{\text{a.e.}}{=} \mathbf{s}_i,\ \xi[\mathbf{u}]\big|_{\mathscr{S}_j + v} \overset{\text{a.e.}}{=} \sigma^v \circ \mathbf{s}_j \right\} \tag{A.44}$$

is the intrinsic viability constraint *of ω, given some fixed RealLife rule. The* boundary *of the constraint is defined:*

$$\partial\mathcal{V}_\omega := \bigcup_{S \in \mathcal{S}} \left\{ \mathbf{u}\big|_{\mathcal{E}^2(\mathscr{S})} \in \mathcal{V}_\omega : \mathbf{u}\big|_{\mathscr{S}} \overset{\text{a.e.}}{=} \mathbf{s},\ \nexists (S, S', v) \in \mathcal{T} \left(\xi[\mathbf{u}]\big|_{\mathscr{S}' + v} \overset{\text{a.e.}}{=} \sigma^v \circ \mathbf{s}' \right) \right\} \tag{A.45}$$

Density Formulation

First, we define a function that gives constraints Ψ given a source and target function:

$$\Gamma[\mathbf{u}_t, \mathbf{u}_{t+1}](x) = \begin{cases} [b_0, b_1] & \mathbf{u}_t(x) = 0 \wedge \mathbf{u}_{t+1} = 1 \\ [0,1] \backslash [b_0, b_1] & \mathbf{u}_t(x) = 0 \wedge \mathbf{u}_{t+1} = 0 \\ [s_0, s_1] & \mathbf{u}_t(x) = 1 \wedge \mathbf{u}_{t+1} = 1 \\ [0,1] \backslash [s_0, s_1] & \mathbf{u}_t(x) = 1 \wedge \mathbf{u}_{t+1} = 0. \end{cases} \tag{A.46}$$

We will also want a way to represent the set of density functions that satisfy a given constraint:

$$\mathcal{D}[\Gamma](\mathscr{A}) := \left\{ \Psi\big|_{\mathscr{A}} : \Psi(x) \in \Gamma(x)\ \forall\, x \in \mathscr{A} \right\}. \tag{A.47}$$

Finally, we need the multidimensional Fourier transform and its inverse for functions in $\mathcal{L}^1(\mathbb{R}^D)$ (i.e., absolutely integrable functions: $\int_{\mathbb{R}^D} |f| \leq \infty$):

$$\mathcal{F}[f](x) := \int_{\mathbb{R}^D} f(y) e^{-i\langle x \cdot y \rangle}\, \lambda[dy] \tag{A.48}$$

$$\mathcal{F}^{-1}[f](x) := \lim_{\epsilon \to 0} \frac{1}{(2\pi)^D} \int_{\mathbb{R}^D} f(y) e^{i\langle x \cdot y \rangle} \zeta_\epsilon(y) \prod_{j=1}^{n} dy_j \tag{A.49}$$

where $\langle x \cdot y \rangle$ denotes the inner product $\sum_{j=1}^{n} x_j y_j$ and $\zeta_\epsilon(y) := \exp(-\epsilon \|y\|_2^2)$ is a cutoff function (Folland, 1992, p. 244).

Now we can state the following conjecture:

Conjecture A.5 (Density Formulation of Viability)**.** Let ω be an autopoietic process dependency graphon by Definition A.12 with an interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ (Definition A.13; Conjecture A.3). Let

$$\mathcal{D}^{ij} := \mathcal{D}[\Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j]](\mathscr{S}_i \cap (\mathscr{S}_j + v)) \tag{A.50}$$

$$U[f] := \mathcal{F}^{-1}\left[\frac{\mathcal{F}[f]}{\mathcal{F}[\kappa]}\right]. \tag{A.51}$$

Then, given a fixed RealLife rule,

$$\mathcal{V}'_\omega := \bigcup_{(S_i, S_j, v) \in \mathcal{T}} \bigcup_{\Psi \in \mathcal{D}^{ij}} \left\{ \mathbf{u}\big|_{\mathcal{E}(\mathscr{S}_j + v)} \overset{\text{a.e.}}{=} U[\Psi'] : \Psi' \in \mathcal{D}[\Gamma[U[\Psi], \sigma^v \circ \mathbf{s}_j]](\mathscr{S}_j + v) \right\} \tag{A.52}$$

is equivalent to the viability constraint defined in Definition A.14 when each $\mathbf{u} \in \mathcal{V}'$ is taken to imply a set of functions over $\mathcal{E}^2(\mathscr{S}_i + v)$ equal over the domain of $\mathbf{u}$.

The titular point of this conjecture is U, which is (almost) bijection from density functions over some set $\mathscr{A}$ to universe state function over $\mathcal{E}(\mathscr{A})$. However, establishing that such a bijection exists is non-trivial. Fortunately, we can use on results from Fourier analysis to get us most of the way. Specifically, we need to use the Convolution Theorem and the Fourier Inverse Theorem. The Convolution Theorem states that, for any two functions $f, g \in \mathcal{L}^1$:

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]. \tag{A.53}$$

This clearly applies to κ and $\mathbf{u} \in \mathcal{U}$, as both are in $\mathcal{L}^1$, assuming that $\mathbf{u}$ has finite support (Folland, 1992, Theorem 7.8). The Fourier Inverse Theorem states that, for any $f \in \mathcal{L}^1$:

$$f \overset{\text{a.e.}}{=} \mathcal{F}^{-1}[\mathcal{F}[f]] \tag{A.54}$$

(Folland, 1992, p. 244). Thus, we get

$$\mathcal{F}[\kappa * \mathbf{u}] = \mathcal{F}[\kappa]\mathcal{F}[\mathbf{u}], \tag{A.55}$$

which by simple algebra and an inverse Fourier transform will give Equation A.51. There is one problem with this claim, however: the Fourier transforms given above are defined over $\mathbb{R}^D$, but in our application here, we are dealing with functions over finite domains ($\mathbf{u}$ over $\mathcal{E}(\mathscr{A})$ and Ψ over $\mathscr{A}$). While we can assume these functions are are zero everywhere outside of their domain, this changes their Fourier transforms, which is potentially problematic for Equation A.51. We therefore require that any extensions of Ψ beyond its domain in $\mathscr{A}$ results in the same U restricted to $\mathcal{E}(\mathscr{A})$. Stated more rigorously, I make the following general claim:

Claim. Let X and Y be subsets of $\mathbb{R}^D$ with non-zero Lebesgue measure, and let $\widetilde{XY} := \cup_{x\in X} Y + x$. Let $g \in \mathcal{L}^1(\mathbb{R}^D)$ be function such that $g(y) = 0$ for almost all $y \notin Y$. Then for every $f, f' \in \mathcal{L}^1(\mathbb{R}^D)$ such that $f(x) \overset{\text{a.e.}}{=} f'(x)$ over X,

$$\mathcal{F}^{-1}[\mathcal{F}[f]/\mathcal{F}[g]](\tilde{x}) = \mathcal{F}^{-1}[\mathcal{F}[f']/\mathcal{F}[g]](\tilde{x}), \tag{A.56}$$

for almost all $\tilde{x} \in \widetilde{XY}$.

Unfortunately, I am not aware of any proofs for this specific claim, but given that $\Psi = \kappa * \mathbf{u}$ can only be defined over $\mathscr{A}$ if $\mathbf{u}$ is defined over $\mathcal{E}(\mathscr{A})$, and that the inverse relation holds in the infinite domain case, it does not appear that this claim is unreasonable.

The density formulation allows us to define two metrics of density functions with respect to a process dependency graphon ω.

Definition A.15 (Transition Metric)**.** *The* transition metric*, assigning a value to how well a density function Ψ satisfies the density bounds implied by the structural transition from $\mathbf{s}_i$ to $\sigma^v \circ \mathbf{s}_j$, is defined:*

$$d_t[\Psi](x) := \inf\left\{|\Psi(x) - y| : y \in \Gamma[\mathbf{s}_i, \sigma^v \circ \mathbf{s}_j](x)\right\}. \tag{A.57}$$

Note that Γ may in part depends on Ψ, by virtue of Ψ being used to extend $\mathbf{s}_i$ over all of $\mathscr{S}_j$. One could thus use d_t over $\mathscr{S}_i \cap \mathscr{S}_j$ with a fixed Γ, and then extend the metric over $\mathscr{S}_j$.

Definition A.16 (Viability Metric)**.** *The* viability metric*, assigning a value to how close a density function is to corresponding to any viable structural transition from S_i, is defined:*

$$d_{\mathcal{V}}[S_i, \Psi](x) := \min\{d_t[\Psi](x) : t = (S_i, \cdot, \cdot) \in \mathcal{T}\} \tag{A.58}$$

Much like the previous metric, Γ can only be fixed over some intersection with $\mathscr{S}_i$. And again, a choice of Ψ can extend the metric over a larger region. Note that $d_{\mathcal{V}} = 0$ does not imply $d_t = 0$, since the parts of $d_{\mathcal{V}}$ that are satisfied may correspond to different Γ. Thus, $d_{\mathcal{V}} = 0$ does not mean that a perturbation is actually viable, but that it satisfies the bounds of at least one transition at every point.

A.2.4. Rule Derivations

Section A.2.3 assumed a fixed RealLife rule $(\rho, b_0, b_1, s_0, s_1)$. However, it is possible to derive a set of rules under which an organisation can be realised, assuming it is reversible and all perturbations in the interaction graph correspond to an empty universe.

The set of rules can be implicitly specified by an 8-tuple $(b_0^L, b_0^U, b_1^L, b_0^U, s_0^L, s_1^U)$, where

$$b_0^L < b_0 \le b_0^U \tag{A.59}$$

$$b_1^L \le b_1 < b_1^U \tag{A.60}$$

$$s_0^L < s_0 \le s_0^U \tag{A.61}$$

$$s_1^L \le s_1 < s_1^U \tag{A.62}$$

$$0 < s_0 \le b_0 < b_1 \le s_1 \le 1. \tag{A.63}$$

To get values for these parameters, we first partition the points in $\mathscr{S}_i \cap (\mathscr{S}_j + v)$ for each $t = (S_i, S_j, v) \in \mathcal{T}$ into four sets:

$$\mathcal{X}_t^b := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 0 \wedge \mathbf{s}_j(x) = 1\} \tag{A.64}$$

$$\mathcal{X}_t^{\neg b} := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 0 \wedge \mathbf{s}_j(x) = 0\} \tag{A.65}$$

$$\mathcal{X}_t^s := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 1 \wedge \mathbf{s}_j(x) = 1\} \tag{A.66}$$

$$\mathcal{X}_t^{\neg s} := \{x \in \mathscr{S}_i \cap (\mathscr{S}_j + v) : \mathbf{s}_i(x) = 1 \wedge \mathbf{s}_j(x) = 0\}. \tag{A.67}$$

I am assuming here that, for every $x \in \Omega$, $W_{c_1}(x,y)$ is either constant or undefined over all $y \in \Omega$. This is implied by the interaction graph property discussed above.

Since I am only considering vacuum perturbations, we can define Ψ over all of $\mathscr{S}$ by assuming that $\mathbf{s}(x) = 0$ for all $x \notin \mathscr{S}$. Then the parameters are defined

$$b_0^U := \min\left\{\inf\left\{\Psi(x) : x \in \mathcal{X}_t^b\right\}, t \in \mathcal{T}\right\} \tag{A.68}$$

$$b_0^L := \max\left\{\sup\left\{\Psi(x) < b_0^U : x \in \mathcal{X}_t^{\neg b}\right\}, t \in \mathcal{T}\right\} \tag{A.69}$$

$$b_1^L := \max\left\{\sup\left\{\Psi(x) : x \in \mathcal{X}_t^b\right\}, t \in \mathcal{T}\right\} \tag{A.70}$$

$$b_1^U := \min\left\{\inf\left\{\Psi(x) > b_1^L : x \in \mathcal{X}_t^{\neg b}\right\}, t \in \mathcal{T}\right\} \tag{A.71}$$

$$s_0^U := \min\left\{\inf\left\{\Psi(x) : x \in \mathcal{X}_t^s\right\}, t \in \mathcal{T}\right\} \tag{A.72}$$

$$s_0^L := \max\left\{\sup\left\{\Psi(x) < s_0^U : x \in \mathcal{X}_t^{\neg s}\right\}, t \in \mathcal{T}\right\} \tag{A.73}$$

$$s_1^L := \max\left\{\sup\left\{\Psi(x) : x \in \mathcal{X}_t^s\right\}, t \in \mathcal{T}\right\} \tag{A.74}$$

$$s_1^U := \min\left\{\inf\left\{\Psi(x) > s_1^L : x \in \mathcal{X}_t^{\neg s}\right\}, t \in \mathcal{T}\right\} \tag{A.75}$$

$$\tag{A.76}$$

These definitions need some qualifications, however. If any of the values do not exist (i.e., when one of $\mathcal{X}$ is empty), they can be ignored in the above inequalities. If such a parameter is needed to compute another one, more care is needed. When $\mathcal{X}^b$ is empty, ignore the b_1^L constraint in the definition of b_0^L; b_1^U can be ignored entirely. Otherwise, one can safely ignore the constraint in a parameter definition when that constraint is undefined.

The logic of this is grounded in the following. First, $\mathcal{X}^{\neg b}$ must always contain a partial process that produces a 0-component due to being sufficiently *below* the b_0 boundary. This is always true since all viable configurations have ρ-wide boundaries; thus, in a vacuum, $\Psi = 0$ for all processes in $\partial\mathscr{S}$. This then means that b_0^L must always exist. Second, since Ψ is continuous with respect to space (the convolution of two bounded functions is always continuous), $\Psi > b_1$ only if $\mathcal{X}_t^b$ is non-empty: there must be an intermediate Ψ between those above and below $[b_0, b_1]$. Finally, I assume that $\mathcal{X}_t^s$ is always non-empty.

This procedure should only derive rules that support the realisation of an organisation as some periodic LtL pattern. I conjecture that this procedure will derive all and only those RealLife rules that support the realisation of a reversible organisation in a vacuum universe. Stated more formally:

Conjecture A.6 (Sufficiency of the RealLife Rule Derivation)**.** Let ω be an autopoietic graphon whose interaction graph $\mathcal{I}_\omega = (\mathcal{S}, \mathcal{T}, \mathfrak{s}, \mathfrak{t}, \tau)$ implies a reversible dynamics: each vertex has exactly one outgoing edge and exactly one incoming edge.

I. All rules that satisfy the constraints derived by the procedure just described are such that, for every $(S_i, S_j, v) \in \mathcal{T}$:

$$\xi[\mathbf{s}_i] \overset{\text{a.e.}}{=} \sigma^v \circ \mathbf{s}_j \tag{A.77}$$

where $\mathbf{s}(x) = 0$ for all $x \notin \mathscr{S}$.

II. The set of rules that satisfy the derived constraints contains *every* rule such that, for every $(S_i, S_j, v) \in \mathcal{T}$, Equation A.77 is true.

A.3. Larger than Life as an Approximation to RealLife

Pivato (2007, Theorem 2.1) proved that large-radius LtL automata are good approximations to RealLife. This section lays out some of the machinery needed to show that process dependency graphons (Definition A.11) are an appropriate limiting object for sequences of increasing large process dependency graphs (Definition A.3). Thus, this section shows that the autopoietic theory developed in LtL is a good approximate theory for autopoiesis in RealLife.

This section depends on results in Graph Limit Theory (Boeckner, 2013; Lovász, 2012; Lovász & Szegedy, 2010) for the converge of dense graph sequences. We can quickly verify that, at least for the block, process dependency graphs are dense. The density of a digraph $G = (V, E)$ can be computed as

$$\frac{|E|}{|V|(|V|-1)}. \tag{A.78}$$

We can attain simple formulas for the number of vertices and edges in ρ^* block organisations:

$$|V(G_{\rho *})| = (3\rho^* + 1)^2 \tag{A.79}$$

$$|E(G_{\rho^*})| = (1 + 4\rho^* + 5(\rho^*)^2)^2. \tag{A.80}$$

We can then compute the limit of the density of block organisations:

$$\lim_{\rho^* \to \infty} \frac{(1 + 4\rho^* + 5(\rho^*)^2)^2}{(3\rho^* + 1)^2((3\rho^* + 1)^2 - 1)} = \frac{25}{81}. \tag{A.81}$$

Thus, block organisations are dense.

To show the converge of process dependency graphs to graphons, we first need to represent process dependency graphs using a RealLife neighbourhood. If G is a K^*-decorated digraph of ρ^*, it can be converted to a K-decorated digraph by scaling the neighbourhood relations of every edge by $\frac{\rho}{\rho^*}$, where ρ is a RealLife neighbourhood radius. That is, for every edge decoration in the graph, $(r, \cdot, \cdot) \mapsto \left(r\frac{\rho}{\rho^*}, \cdot, \cdot\right)$.

Let $\mathcal{Q}$ denote the set of functions $f : K \to \mathbb{R}$. Let $E_G^d(ij)$ give the decoration the graph G assigns to the edge ij. Then the homomorphism density of a $\mathcal{Q}$-decorated graph F into a K-decorated graph G_{ρ^*} is defined

$$t(F, G_{\rho *}) := \frac{1}{V(G_{\rho^*})^{V(F)}} \sum_{f:V(F)\to V(G_{\rho^*})} \prod_{ij\in E(F)} E_F^d(ij) \left(E_{G_{\rho^*}}^d(f(i)f(j))\right). \tag{A.82}$$

The homomorphism density of a $\mathcal{Q}$-decorated graph F into a process dependency graphon ω can also be defined:

$$t(F, \omega) := \lambda \left[\Omega^{V(F)}\right]^{-1} \int_{\Omega^{V(F)}} \prod_{ij\in E(F)} \left[\int_{K_0} E_F^d(ij)(k)\omega(x_i, x_j)(dk)\right] \lambda[dx]. \tag{A.83}$$

A sequence of graphs (G_{ρ^*}) converges to the graphon ω if

$$\lim_{\rho^* \to \infty} t(F, G_{\rho^*}) = t(F, \omega). \tag{A.84}$$

for every $\mathcal{Q}$-decorated graph F.

Equation A.83 can be simplified with an appropriate choice of $\mathcal{R} \subset \mathcal{Q}$ and g. Let $W_f(x,y) := \int_{K_0} f(k)\,\omega(x,y)\,(dk)$. Then the homomorphism density of a $\mathcal{R}$-decorated graph F into a graphon ω is defined:

$$t(F,\omega) = \lambda\left[\Omega^{V(F)}\right]^{-1} \int_{\Omega^{V(F)}} \prod_{ij \in E(F)} \left[W_{E^d_F(ij)}(x_i,x_j)\right] \lambda[dx]. \tag{A.85}$$

If $\mathcal{R}$ is a *generating system* of $\mathcal{Q}$ ($\mathcal{R}$ is countable and every $f \in \mathcal{Q}$ can be expressed as a linear combination of functions in $\mathcal{R}$) and $W_f(x,y)$ is (i) linear in $f \in \mathcal{Q}$, and (ii) $W_f \in [\inf\{f^{\rightarrow}\}, \sup\{f^{\rightarrow}\}]$, for every fixed $(x,y) \in \Omega^2$, then the Riesz representation theorem guarantees the existence of a graphon ω such that $W_f(x,y) := \int_{K_0} f(k)\,\omega(x,y)\,(dk)$, and convergence for every $f \in \mathcal{R}$ is equivalent to convergence for every $f \in \mathcal{Q}$ (Lovász, 2012, p. 323). This means that we only need to know values of W_f for every $f \in \mathcal{R}$ to prove convergence of homomorphism densities for every fixed $\mathcal{Q}$-decorated graph.

The examples of $\mathcal{R}$ provided in Chapter 3 do not satisfy all of these criteria. Specifically, they are not generating systems of $\mathcal{Q}$. They do however, fulfill the other criteria. Thus, they demonstrate that block sequences satisfy at least some conditions of convergence (that is, convergence for a subset of $\mathcal{Q}$).

Bibliography

Adamatzky, A. (2010). *Game of Life cellular automata*. Springer-Verlag. https://doi.org/10.1007/978-1-84996-217-9

Agmon, E., Gates, A. J., Churavy, V., & Beer, R. D. (2016). Exploring the space of viable configurations in a model of metabolism–boundary co-construction. *Artificial Life*, *22*(2), 153–171. https://doi.org/10.1162/ARTL_a_00196

Alon, U. (2019). *An introduction to systems biology: Design principles of biological circuits* (2nd ed.). Chapman Hall/CRC. https://doi.org/10.1201/9780429283321

Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour* (2nd ed.). Springer Dordrecht. https://doi.org/10.1007/978-94-015-1320-3

Aubin, J. P. (1991). *Viability theory*. Springer Science+Busniness Media.

Ball, P. (2023). *How life works: A user's guide to the new biology*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226826691

Barandiaran, X. E., & Egbert, M. D. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, *20*(1), 5–28. https://doi.org/10.1162/ARTL_a_00094

Barrett, N. F. (2017). The normative turn in enactive theory: An examination of its roots and implications. *Topoi*, *36*(3), 431–443. https://doi.org/10.1007/s11245-015-9355-y

Beaujean, P., Sikora, F., & Yger, F. (2021). Scaling up graph homomorphism features with efficient data structures. *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*. https://openreview.net/forum?id=EwT8NpZIth8

Beer, R. D. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*(1–2), 173–215. https://doi.org/10.1016/0004-3702(94)00005-L

Beer, R. D. (1995b). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, *3*(4), 469–509. https://doi.org/10.1177/105971239500300405

Beer, R. D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems*, *20*(2–4), 257–289. https://doi.org/10.1016/S0921-8890(96)00063-2

Beer, R. D. (2004). Autopoiesis and cognition in the Game of Life. *Artificial Life*, *10*(3), 309–326. https://doi.org/10.1162/1064546041255539

Beer, R. D. (2008). The dynamics of brain–body–environment systems: A status report. *Handbook of Cognitive Science*, 99–120. https://doi.org/10.1016/B978-0-08-046616-3.X0001-1

Beer, R. D. (2014). The cognitive domain of a glider in the Game of Life. *Artificial Life*, *20*(2), 183–206. https://doi.org/10.1162/ARTL_a_00125

Beer, R. D. (2015). Characterizing autopoiesis in the Game of Life. *Artificial Life*, *21*(1), 1–19. https://doi.org/10.1162/ARTL_a_00143

Beer, R. D. (2020a). Bittorio revisited: Structural coupling in the Game of Life. *Adaptive Behavior*, *28*(4), 197–212. https://doi.org/10.1177/1059712319859907

Beer, R. D. (2020b). An integrated perspective on the constitutive and interactive dimensions of autonomy. In J. Bongard, J. Lovato, L. Herbert-Dufrésne, R. Dasari, & L. Soros (Eds.), *ALIFE 2020: The 2020 conference on artificial life* (pp. 202–209). MIT Press. https://doi.org/10.1162/isal_a_00245

Beer, R. D. (2020c). An investigation into the origin of autopoiesis. *Artificial Life*, *26*(1), 5–22. https://doi.org/10.1162/artl_a_00307

Beer, R. D. (2020d). Lost in words. *Adaptive Behavior*, *28*(1), 19–21. https://doi.org/10.1177/1059712319867907

Beer, R. D. (2023). On the proper treatment of dynamics in cognitive science. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12686

Beer, R. D., Barwich, A.-S., & Severino, G. J. (2024). Milking a spherical cow: Toy models in neuroscience. *European Journal of Neuroscience*, *60*(10), 6359–6374. https://doi.org/10.1111/ejn.16529

Beer, R. D., & Di Paolo, E. A. (2023). The theoretical foundations of enaction: Precariousness. *BioSystems*, *223*, 104823. https://doi.org/10.1016/j.biosystems.2022.104823

Beer, R. D., McShaffrey, C., & Gaul, T. M. (2024). Deriving the intrinsic viability constraint of an emergent individual from first principles. In A. Faíña, S. Risi, E. Medvet, K. Stoy, B. Chan, K. Miras, P. Zahadat, D. Grbic, & G. Nadizar (Eds.), *ALIFE 2024: Proceedings of the 2024 artificial life conference* (pp. 192–200). MIT Press. https://doi.org/10.1162/isal_a_00739

Berlekamp, E. R., Conway, J. H., & Guy, R. K. (1982). What is Life? In E. R. Berlekamp, J. H. Conway, & R. K. Guy (Eds.), *Winning ways for your mathematical plays* (pp. 817–850, Vol. 2). Academic Press. https://doi.org/10.1080/00029890.1986.11971843

Boden, M. A. (2000). Autopoiesis and life. *Cognitive Science Quarterly*, *1*(1), 115–143.

Boeckner, D. (2013). *Directed graph limits and directed threshold graphs* [Doctoral dissertation, University of Nebraska].

Bourgine, P., & Stewart, J. (2004). Autopoiesis and cognition. *Artificial life*, *10*(3), 327–345. https://doi.org/10.1162/1064546041255557

Bourgine, P., & Varela, F. J. (1992). Towards a practice of autonomous systems. In P. Bourgine & F. J. Varela (Eds.), *Toward a practice of autonomous systems: Proceedings of the first European conference on artificial life* (pp. xi–xvii, Vol. 1). MIT Press.

Colón-Santos, S., Vázquez-Salazar, A., Adams, A., Campillo-Balderas, J. A., Hernández-Morales, R., Jácome, R., Muñoz-Velasco, I., Rodriguez, L. E., Schaible, M. J., Schaible, G. A., et al. (2024). Chapter 2: What is life? *Astrobiology*, *24*(S1), S-40–S-56. https://doi.org/10.1089/ast.2021.0116

Cornish-Bowden, A. (2006). Putting the systems back into systems biology. *Perspectives in Biology and Medicine*, *49*(4), 475–489. https://doi.org/10.1353/pbm.2006.0053

Cornish-Bowden, A. (2011). How far has it come?: Systems biology. *The Biochemist*, *33*(1), 16–18. https://doi.org/10.1042/BIO03301016

Cornish-Bowden, A., & Cárdenas, M. L. (2020). Contrasting theories of life: Historical context, current theories. in search of an ideal theory. *BioSystems*, *188*, 104063. https://doi.org/10.1016/j.biosystems.2019.104063

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452. https://doi.org/10.1007/s11097-005-9002-y

Di Paolo, E. A. (2009). Extended life. *Topoi*, *28*, 9–21. https://doi.org/10.1007/s11245-008-9042-3

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT Press. https://doi.org/10.7551/mitpress/11244.001.0001

Egbert, M. D., & Di Paolo, E. (2009). Integrating autopoiesis and behavior: An exploration in computational chemo-ethology. *Adaptive Behavior*, *17*(5), 387–401. https://doi.org/10.1177/1059712309343821

Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organization*. Springer-Verlag. https://doi.org/10.1007/978-3-642-67247-7

Evans, K. M. (1996). *Larger than Life: It's so nonlinear* [Doctoral dissertation, The University of Wisconsin-Madison].

Evans, K. M. (2001). Larger than Life: Digital creatures in a family of two-dimensional cellular automata. *Discrete Mathematics and Theoretical Computer Science*, *AA*, 177–192. https://doi.org/10.46298/dmtcs.2288

Evans, K. M. (2003). Larger than Life: Threshold-range scaling of Life's coherent structures. *Physica D: Nonlinear Phenomena*, *183*(1-2), 45–67. https://doi.org/10.1016/S0167-2789(03)00155-6

Evans, K. M. (2009). Larger than Life's invariant measures. *Electronic Notes in Theoretical Computer Science*, *252*, 55–75. https://doi.org/10.1016/j.entcs.2009.09.014

Favela, L. H. (2020). Dynamical systems theory in cognitive science and neuroscience. *Philosophy Compass*, *15*(8), e12695. https://doi.org/10.1111/phc3.12695

Fleischaker, G. R. (1988). Autopoiesis: The status of its system logic. *BioSystems*, *22*(1), 37–49.

Folland, G. B. (1992). *Fourier analysis and its applications*. Wadsworth & Brooks/Cole.

Fontana, W., & Buss, L. W. (1994). "The arrival of the fittest": Toward a theory of biological organization. *Bulletin of Mathematical Biology*, *56*, 1–64. https://doi.org/10.1007/BF02458289

Foucault, M. (1975). *The order of things*. Routledge. https://doi.org/10.4324/9780203996645

Frisch, U., Hasslacher, B., & Pomeau, Y. (1986). Lattice-gas automata for the Navier-Stokes equation. *Physical Review Letters*, *56*(14), 1505.

Gánti, T. (2003). *The principles of life*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198507260.001.0001

Garfinkel, A., Shevtsov, J., & Guo, Y. (2017). *Modeling life: The mathematics of biological systems*. Springer Cham. https://doi.org/10.1007/978-3-319-59731-7

Gaul, T. M. (2024). Autopoiesis in RealLife Euclidean automata. In A. Faíña, S. Risi, E. Medvet, K. Stoy, B. Chan, K. Miras, P. Zahadat, D. Grbic, & G. Nadizar (Eds.), *ALIFE 2024: Proceedings of the 2024 artificial life conference* (pp. 818–826). MIT Press. https://doi.org/10.1162/isal_a_00723

Gaul, T. M., & Izquierdo, E. J. (2025). *Cognitive distinctions as a language for cognitive science: Comparing methods of description in a model of referential communication* [To Appear in *Artificial Life*].

Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M., & Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. *PLOS Computational Biology*, *14*(2), e1005991. https://doi.org/10.1371/journal.pcbi.1005991

Hodgins, G. W. (2009). Measuring atomic bomb-derived 14c levels in human remains to determine year of birth and/or year of death. *NIJ Report*, *98*.

Jacob, F. (1982). *The possible and the actual*. University of Washington Press.

Jonas, H. (2001). *The phenomenon of life: Toward a philosophical biology*. Northwestern University Press.

Judson, H. F. (2013). *The eighth day of creation: The makers of the revolution in biology* (Commemorative). CSHL Press.

Kant, I. (2000). *Critique of the power judgment* (P. Guyer, Ed.; P. Guyer & E. Mattews, Trans.). Cambridge University Press. https://doi.org/10.1017/CBO9780511804656 (Original work published 1790)

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., & Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, *150*(2), 389–401. https://doi.org/10.1016/j.cell.2012.05.044

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, *119*(1), 1–24. https://doi.org/10.1016/S0022-5193(86)80047-9

Kirschner, M., Gerhart, J., & Mitchison, T. (2000). Molecular ”vitalism”. *Cell*, *100*(1), 79–88. https://doi.org/10.1016/S0092-8674(00)81685-2

Krüger, T., Halim, K., Kuzmin, A., Shardt, O., Silva, G., & Viggen, E. M. (2017). *The lattice boltzmann method: Principles and practice*. Springer Cham. https://doi.org/10.1007/978-3-319-44649-3

Lovász, L. (2012). *Large networks and graph limits* (Vol. 60). American Mathematical Society Colloquium Publications.

Lovász, L., & Szegedy, B. (2010). Limits of compact decorated graphs. *arXiv preprint arXiv:1010.5155*.

Luisi, P. L. (2003). Autopoiesis: A review and a reappraisal. *Naturwissenschaften*, *90*, 49–59. https://doi.org/10.1007/s00114-002-0389-9

Margulis, L., & Sagan, D. (2000). *What is life?* University of California Press.

Maturana, H. R. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, *9*(3-4), 5–34.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Springer Netherlands.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.

Mayr, E. (1961). Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science*, *134*(3489), 1501–1506. https://doi.org/10.1126/science.134.3489.1501

McMullin, B. (1999). Some remarks on autocatalysis and autopoiesis. *Closure: Emergent Organizations and Their Dynamics*. https://doi.org/10.1111/j.1749-6632.2000.tb06276.x

McMullin, B. (2004). Thirty years of computational autopoiesis: A review. *Artificial Life*, *10*(3), 277–295. https://doi.org/10.1162/1064546041255548

McShaffrey, C., & Beer, R. D. (2022). Maintaining viability with multiple needs. *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1523–1528. https://doi.org/10.1109/SSCI51031.2022.10022260

McShaffrey, C., & Beer, R. D. (2023). Decomposing viability space. In H. Iizuka, K. Suzuki, R. Uno, L. Damiano, N. Spychala, M. Aguilera, E. Izquierdo, R. Suzuki, & M. Baltieri (Eds.), *ALIFE 2023: Ghost in the machine: Proceedings of the 2023 artificial life conference* (pp. 357–364). MIT Press. https://doi.org/10.1162/isal_a_00652

Merleau-Ponty, M. (1963). *The structure of behavior* (A. L. Fisher, Trans.). Beacon Press. (Original work published 1942)

Merleau-Ponty, M. (2010). *Phenomenology of perception* (C. Lefort, T. Carman, & D. Landes, Eds.; D. Landes, Trans.). Routledge. https://doi.org/10.4324/9780203720714 (Original work published 1945)

Monod, J. (1971). *Chance and necessity: An essay on the natural philosophy of modern biology*. Alfred A. Knopf.

Montévil, M., & Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of Theoretical Biology*, *372*, 179–191. https://doi.org/10.1016/j.jtbi.2015.02.029

Morange, M. (2005). What is Life? a new look at an old question. *Logic, Methodology and Philosophy of Science*, 431–440.

Morange, M. (2010). The resurrection of life. *Origins of Life and Evolution of Biospheres*, *40*(2), 179–182. https://doi.org/10.1007/s11084-010-9197-y

Morange, M. (2021). *A history of biology*. Princeton University Press. https://doi.org/10.1515/9780691188782

Mukherjee, S. (2022). *The song of the cell: An exploration of medicine and the new human*. Simon & Schuster.

Nurse, P. (2021). *What is life?: Five great ideas in biology*. W.W. Norton & Company.

Ono, N., & Ikegami, T. (2000). Self-maintenance and self-reproduction in an abstract cell model. *Journal of Theoretical Biology*, *206*(2), 243–253. https://doi.org/10.1006/jtbi.2000.2121

Oppenheim, A. V., & Schafer, R. W. (2010). *Discrete-time signal processing* (3rd ed.). Pearson.

Pivato, M. (2007). RealLife: The continuum limit of Larger than Life cellular automata. *Theoretical Computer Science*, *372*(1), 46–68. https://doi.org/10.1016/j.tcs.2006.11.019

Pross, A. (2016). *What is life?: How chemistry becomes biology*. Oxford University Press.

Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., & Krakauer, D. C. (2022). *Protocells: Bridging nonliving and living matter*. MIT Press. https://doi.org/10.7551/mitpress/9780262182683.001.0001

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. Colubmia University Press.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, *10*(3), 235–259. https://doi.org/10.1162/1064546041255584

Smolin, L. (2004). Cosmological natural selection as the explanation for the complexity of the universe. *Physica A: Statistical Mechanics and its Applications*, *340*(4), 705–713. https://doi.org/10.1016/j.physa.2004.05.021

Sourjik, V., & Wingreen, N. S. (2012). Responding to chemical gradients: Bacterial chemotaxis. *Current Opinion in Cell Biology*, *24*(2), 262–268. https://doi.org/10.1016/j.ceb.2011.11.008

Strogatz, S. H. (2024). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering* (Third). CRC Press.

Sun, G., Ahn-Horst, T. A., & Covert, M. W. (2021). The E. coli whole-cell modeling project. *EcoSal plus*, *9*(2), eESP–0001. https://doi.org/10.1128/ecosalplus.ESP-0001-2020

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press. https://doi.org/10.7551/mitpress/2524.001.0001

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Varela, F. J. (1979). *Principles of biological autonomy*. North Holland.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and cognition*, *34*(1), 72–87.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5*(4), 187–196. https://doi.org/10.1016/0303-2647(74)90031-8

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind, revised edition: Cognitive science and human experience* (Revised). MIT Press. https://doi.org/10.7551/mitpress/9780262529365.001.0001

Virgo, N., Egbert, M. D., & Froese, T. (2011). The role of the spatial boundary in autopoiesis. In G. Kampis, I. Karsai, & E. Szathmáry (Eds.), *Advances in artificial life. darwin meets von neumann* (pp. 240–247). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21283-3_30

von Uexküll, J. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds (C. H. Schille, Trans.). *Semiotica*, *89*(4), 319–391. https://doi.org/10.1515/semi.1992.89.4.319 (Original work published 1934)

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*(2), 97–125. https://doi.org/10.1023/A:1020368120174

Wilmsen, S., & Kost, C. (2025). Defining organismality. *Biological Theory*, *20*, 54–70. https://doi.org/10.1007/s13752-024-00486-0

Wittgenstein, L. (2009). *Philosophical investigations* (P. M. S. Hacker & J. Shulte, Eds.; G. E. M. Anscombe, P. M. S. Hacker, & J. Shulte, Trans.; Fourth). John Wiley & Sons. (Original work published 1953)

Woese, C. R. (2004). A new biology for a new century. *Microbiology and Molecular Biology Reviews*, *68*(2), 173–186. https://doi.org/10.1128/mmbr.68.2.173-186.2004

Wolf-Gladrow, D. A. (2000). *Lattice-gas cellular automata and lattice Boltzmann models: An introduction*. Springer Berlin, Heidelberg. https://doi.org/10.1007/b72010

CIRRICULUM VITAE

EDUCATION

Indiana University Bloomington, Bloomington, IN *2021 - 2025*
B.S. in Cognitive Science

- Thesis: *Biological Organisation and Viability: A Theory of Autopoiesis in Cellular and Euclidean Automata.* Committee: Dr. Randall D. Beer (Chair), Dr. Luis H. Favela, Dr. Eduardo J. Izquierdo.
- Departmental Honors in Cognitive Science
- Major concentration in Computation
- Dean's List

PUBLICATIONS

1. **Gaul**, **T. M.**, & Izquierdo, E. J. (2025). *Cognitive distinctions as a language for cognitive science: Comparing methods of description in a model of referential communication* [To Appear in *Artificial Life*]
2. Beer, R. D., McShaffrey, C., & **Gaul**, **T. M.** (2024). Deriving the intrinsic viability constraint of an emergent individual from first principles. In A. Faíña, S. Risi, E. Medvet, K. Stoy, B. Chan, K. Miras, P. Zahadat, D. Grbic, & G. Nadizar (Eds.), *ALIFE 2024: Proceedings of the 2024 artificial life conference* (pp. 192–200). MIT Press. https://doi.org/10.1162/isal_a_00739
3. **Gaul**, **T. M.** (2024). Autopoiesis in RealLife Euclidean automata. In A. Faíña, S. Risi, E. Medvet, K. Stoy, B. Chan, K. Miras, P. Zahadat, D. Grbic, & G. Nadizar (Eds.), *ALIFE 2024: Proceedings of the 2024 artificial life conference* (pp. 818–826). MIT Press. https://doi.org/10.1162/isal_a_00723

HONORS AND AWARDS

Outstanding Research Award, *Indiana University Cognitive Science.* *March 2025*

Outstanding Contribution Award, *Indiana University Cognitive Science.* *April 2024*

Robert J. Glushko Award for Outstanding Research in Cognitive Science, *Indiana University Cognitive Science.* *April 2023*

SERVICE

Midwest Undergraduate Cognitive Science Conference *Fall 2023 - Spring 2025*
Organizer

· Outreach, finances, and scheduling

Students of Cognitive Science *Spring 2024 - Spring 2025*
Treasurer

· Finances and scheduling

Academic Fairness Committee *Fall 2023*

Committee Member

- Sat in hearings and voted.

PRESENTATIONS

1. **Gaul, T. M.** “Autopoiesis in RealLife Euclidean automata.” *2024 Artificial Life Conference.* Copenhagen, Denmark, 2024.
2. **Gaul, T. M.** “The biology and cognition of emergent individuals in RealLife Euclidean automata.” *Sixteenth Annual Midwest Undergraduate Cognitive Science Conference.* Indiana University Bloomington, IN, 2024.
3. **Gaul, T. M.** “Communication and embedded autonomy: a minimal model of referential communication.” *Fifteenth Annual Midwest Undergraduate Cognitive Science Conference.* Indiana University Bloomington, IN, 2023.