

Multiscale Structure in Eco-Evolutionary Dynamics

Blake C. Stacey

September 8, 2015

University of Massachusetts Boston
Department of Physics

I've done this because I love the act of writing, which is to say I love the act of discovery, of revelation, and then the attempt to share that revelation in all its fullness and clarity.
You can never be sure how many people will want to share in that feeling.

—[TA-NEHISI COATES](#) (2014)

This is a slightly expanded version of my PhD thesis, which was accepted by Brandeis University's Martin A. Fisher School of Physics in May of 2015. In turn, that thesis was based on research I began at the New England Complex Systems Institute.

Contents

1	Abstract	9
2	Multiscale Structure	11
2.1	Introduction	11
2.1.1	Information-Theoretic Axioms for Structured Systems	13
2.1.2	Indices of Structure	14
2.2	Examples	16
2.2.1	Three-Component Systems	16
2.2.2	Minimal Incidence Geometry	17
2.2.3	Fano Plane	21
2.3	The Dual of a Complex System	23
2.4	Computation and Gammoids	26
2.5	Network Dynamics	30
2.6	Frequency-Dependent Moran Process	32
2.7	Multiscale Challenges and Evolution	35
3	Host–Consumer Evolution by Simulation	39
3.1	Introduction	39
3.2	Model and Methods	41
3.3	Results	44
3.3.1	Evolution of Transmissibility	44
3.3.2	Timescale Dependence of Invasion Success	47
3.3.3	Pair Approximations	49
3.3.4	Organism Swapping	50
3.3.5	Effect of Substrate Topology	53
3.3.6	Percolation	53
3.3.7	Patch Size and Structure	56
3.4	Discussion	64
3.4.1	Defining Fitness	66
3.4.2	Pair Approximations and Stability	70
3.4.3	Percolation	71
3.4.4	Adaptive Networks	71
3.4.5	Conclusions	72
4	A Volunteer’s Dilemma	73
4.1	Well-Mixed Populations with Carrying Capacity	73
4.2	Volunteer’s Dilemma in a Networked Population	78

Contents

4.3	Fully Occupied Networks	80
4.3.1	Initially Balanced Population	82
4.3.2	Invasion	82
4.3.3	Mutation–Selection Equilibrium	83
4.3.4	Analytical Results	83
5	Techniques of Probability	101
5.1	Basic Properties	101
5.2	An Analogy with Evolutionary Dynamics	108
5.3	Biased Coin-Flips and Urn Models	111
5.4	Shannon Information	115
5.5	Moments and Cumulants	119
5.6	Generating Functions	125
5.7	Central Limit Theorem	127
5.8	Variations on Diffusion	131
6	Stochastic Adaptive Dynamics	135
6.1	Justifying the Fokker–Planck Equation	135
6.2	The Deterministic Limit	138
6.3	A Fokker–Planck Equation for Adaptive Dynamics	140
6.4	Concurrent Mutations, Discreteness and Multi-strategy Games	145
6.5	Interspecies Interactions	149
7	Spatial Stochastic Mechanics	153
7.1	The Central Limit Theorem by RG Transformations	153
7.2	Isotropic Percolation	160
7.3	Doi Formalism	165
7.4	Examples	169
7.5	Coherent-State Path Integrals	171
7.6	Spatial Dependence	173
7.7	Rate Equations from Tree-Level Calculations	175
7.8	Directed Percolation	177
7.9	Prior Relevant Results and Difficulties	178
7.10	Carrying Capacity	179
7.11	A Common Denominator	180
8	Invasion Fitness by Moment Closure Approximations	187
8.1	Introduction and Overview	187
8.2	Example 1: Birth, Death, Movement	189
8.3	Example 2: Epidemic in an Adaptive Network	191
8.4	Example 3: Evolution of Altruism	194
8.5	Host–Consumer Dynamics	199
8.6	Modified Mean-Field	201

9	The Varieties of Multilevel Selection	205
9.1	Introduction	205
9.2	Fisher's Fundamental Theorem	209
9.3	The Price Equation: Motivation and Shortcomings	211
9.4	Interconverting MLS-A and Inclusive Fitness	214
9.5	An Example of MLS-B	218
9.6	A Literature of Confusion	220
9.7	Discussion	223
10	Speculations for New Mathematics	225
10.1	More on Multiscale Information	225
10.2	Category Theory for Moment Closures	228
10.3	Games with Variable Numbers of Players	229
10.4	Composition and a Multiscale Doi Formalism	230
10.5	Multiplayer Games and Biodiversity Indices	234
10.6	Gauge Theory and Evolution	243
11	Conclusions	245
11.1	Review	245
11.2	Just One More Thing	249
	Bibliography	251

1 Abstract

In a complex system, the individual components are neither so tightly coupled or correlated that they can all be treated as a single unit, nor so uncorrelated that they can be approximated as independent entities. Instead, patterns of interdependency lead to structure at multiple scales of organization. Evolution excels at producing such complex structures. In turn, the existence of these complex interrelationships within a biological system affects the evolutionary dynamics of that system. I present a mathematical formalism for multiscale structure, grounded in information theory, which makes these intuitions quantitative, and I show how dynamics defined in terms of population genetics or evolutionary game theory can lead to multiscale organization. For complex systems, “more is different,” and I address this from several perspectives. Spatial host–consumer models demonstrate the importance of the structures which can arise due to dynamical pattern formation. Evolutionary game theory reveals the novel effects which can result from multiplayer games, nonlinear payoffs and ecological stochasticity. Replicator dynamics in an environment with mesoscale structure relates to generalized conditionalization rules in probability theory.

The idea of natural selection “acting at multiple levels” has been mathematized in a variety of ways, not all of which are equivalent. We will face down the confusion, using the experience developed over the course of this thesis to clarify the situation.

Chapter 2 applies the general abstract framework of multiscale structure to some geometrical examples, to build intuition for it, and then connects it with population genetics and network theory. Chapter 3 studies emergent multiscale structure in a spatial evolutionary ecosystem. Next, Chapter 4 takes a different approach to the notion of “more is different,” using both simulations and dynamical systems theory to understand evolutionary games in which the interactions do not resolve into pairs.

I have set aside Chapter 5 to summarize the parts of probability theory which will be necessary for the following two chapters, because I’ve yet to find a textbook which has the necessary stuff all in one place. Chapter 6 is purely analytical: I break a theorem from the literature, show how to fix it and then point out where it will break again. The goal for Chapter 7 is to provide analytical arguments for at least a few of the things seen in Chapters 2, 3 and 4. Specifically, I aim to use universality to predict critical exponents for phase transitions.

Chapters 8 and 9 are mostly about explaining other people’s work in a way I can understand. Chapter 10 is essentially a concept piece, intended to sketch out the possibility of new interesting problems.

2 Multiscale Structure

2.1 Introduction

A century and odd years ago, the philosopher William James asked [2],

What shall we call a *thing* anyhow? It seems quite arbitrary, for we carve out everything, just as we carve out constellations, to suit our human purposes. For me, this whole ‘audience’ is one thing, which grows now restless, now attentive. I have no use at present for its individual units, so I don’t consider them. So of an ‘army,’ of a ‘nation.’ But in your own eyes, ladies and gentlemen, to call you ‘audience’ is an accidental way of taking you. The permanently real things for you are your individual persons. To an anatomist, again, those persons are but organisms, and the real things are the organs. Not the organs, so much as their constituent cells, say the histologists; not the cells, but their molecules, say in turn the chemists.

The Jamesian view is that none of these scientific disciplines ought to be taken as more “fundamental” than another. Each must prove its own worth by way of its pragmatic utility; none is by necessity merely the reduction of another to a special case.

In the study of complex systems, we face this directly. A complex system exhibits structure at many scales of organization. For example, one can study human beings at any magnification, from the molecular level to the societal, and an entire science flourishes at each level. We have developed a formalism for making this intuition mathematically precise and quantitatively useful, employing the tools of information theory [3,4,5,6,7,8]. To explore how this formalism can be used, and to make clear the intricacies of multiscale information theory, we shall in this chapter apply that theory to an illustrative class of geometrical problems. Having done this, we will be in a good position to use it to study collective behaviors in systems developed in mathematical biology.

Thinking clearly about what we mean by “complexity” is important for biology, and few problems bring this home more clearly than the so-called *C-value paradox*. This is the puzzle that the sizes of species’ genomes do not correlate with any obvious, intuitive or meaningful measure of organismal complicatedness [9]. A species’ C-value is the characteristic amount of DNA which occurs in one set of chromosomes within its nucleus. It can be measured in picograms, for a physical unit, or in base pairs, for a more informational one. (A trillionth of a gram roughly works out to a billion base pairs.) One might think that species with larger C-values would be more “complex” by some fairly apparent standard. However, nature has not turned out that way. The domestic onion, *Allium cepa*, has approximately 16 billion base pairs on one set of

chromosomes. This is roughly five times the total size of the human genome [10]. And the problem goes beyond the check to our pride: life forms which seem by all accounts to be comparably complicated can have widely separated genome sizes. For example, lungfish can have genomes 350 times larger than those of pufferfish [10]. Even close evolutionary proximity is no guarantee that C-values will agree. *Zea mays*, the maize plant, diverged from the teosinte grass *Zea luxurians* about 140,000 years ago [11], and in that time, its genome size has increased by half [9, 12].

So, the puzzle: what does all that extra genetic information do? The answer, in brief terms, is *basically nothing*.

Nor are we humans making much use of our own genomes, percentage-wise. Eukaryotic DNA contains, as a general rule, vast supplies of *junk* [9, 10, 13, 14, 15, 16, 17, 18, 19]. Some DNA specifies the sequences of proteins, and is designated “coding” DNA. Other stretches of the double helix play a role in regulating which genes are active and when. Still other portions of the genome are transcribed into the RNA components of cellular machinery like ribosomes. But even with all these accounted for, there remain sequences which are, identifiably, detritus.

The C-value “paradox” is not so paradoxical after all, then: the variable amounts of genomic bloat due to nonfunctional DNA make C-value variations a rather unsurprising phenomenon. The presence of a large quantity of nonfunctional DNA can have a biological effect, since it takes up space and increases the resources required for cells to replicate. For example, salamanders carry a truly remarkable amount of genetic information, with different species possessing genomes four to thirty-five times the size of our own [10]. Plainly, it is possible to make a salamander using far less DNA than some species of them have. Among salamander species, larger genome size is correlated with slower regeneration of lost limbs, suggesting that elevated genome size might be somewhat costly [20]. The essential points are, first, that this cost is, if it exists, not on the whole deleterious enough for natural selection to act strongly against it [10], and second, that it is due to the *quantity* of DNA present, not its *specific sequence*.

Moreover, the presence of DNA detritus suggests a way of thinking about complexity more quantitatively. The matter is one of *effective description*. Let us focus, for the moment, on the complexity of a genome itself, rather than of the body plan associated with it. Our intuition leads us to say that we require more information to describe a more complicated genome. However, large stretches of a eukaryotic genome will be junk sequences, which can be switched with other, equally nonfunctional sequences or even deleted entirely with little or no effect. This suggests a strategy: we can describe the functional portion of the genome—the protein-coding genes, the regulatory regions and so forth—faithfully, and then we can loosely characterize the rest. We take careful notes about the functional parts, and then we fill in the rest with broad brush-strokes. A coarse-grained description of the nonfunctional portion is adequate, because any other nucleotide sequence which satisfies the same coarse-grained criteria could be swapped in for the actual junk.

In turn, we can apply the same method to the functional portion. Multiple nucleotide sequences are translated to the same protein, because multiple codons in the genetic code stand for the same amino acid [21, 22]. We can, therefore, exploit this redundancy

and use a smaller number of characters to represent *what it is most important to know* about each protein-coding gene. Then, multiple amino-acid sequences are often biologically equivalent to one another, because the substitution of one amino acid for a similar one does not drastically change the resulting protein [23]. So, we can describe a protein in a coarse-grained way, and so on. Indeed, this plurality of possible sequences compatible with the same coarse-grained description is biologically essential: given the amount of DNA which human cells carry, we would otherwise be ground under heel by the genetic load of deleterious mutations. It also affects the *rate* of evolution, since a population can more easily explore the space of possible genomes when there is a network of neutral paths through it [24].

The general lesson is that *partial descriptions* of a system can have increased utility if they can exploit patterns and redundancies. Furthermore, the way the utility of a description increases as we allow more information to be used tells us about the structure of the system we are describing. This approach is different from the way information theory has typically been used in the past, because we are considering *scale* and information as *complementary quantities* [3]. We measure the effort which goes into a description in units of information, whereas the effectiveness of that description is the scale of what it captures.

We now turn to mathematizing this idea, following earlier work by the author and others on multiscale structure and information theory. The resulting formalism will be applicable at levels from the intracellular to the societal. This will allow us to discuss descriptions, utilities and related concepts beginning from an axiomatic starting point (so that we will not need molecular biology in order to define utility). With these concepts developed and some illustrative examples analyzed, we will then apply them to evolutionary dynamics.

2.1.1 Information-Theoretic Axioms for Structured Systems

For convenience, we review the basic axioms of the multiscale information formalism which we developed in earlier work [3]. In this formalism, a *system* is defined by a set of components, A , and an *information function*, H , which assigns a nonnegative real number to each subset $U \subset A$. This number $H(U)$ is the amount of information needed to describe the components in U . To qualify as an information function, H must satisfy two axioms:

- *Monotonicity*: The information in a subset U that is contained in a subset V cannot have more information than V . That is, $H(U) \leq H(V)$.
- *Strong subadditivity*: Given two subsets, the information contained in both cannot exceed the information in each of them separately minus the information in their intersection:

$$H(U \cup V) \leq H(U) + H(V) - H(U \cap V). \quad (2.1)$$

Given an information function H , we can construct functions which express different kinds of possible correlations among a system's components, such as the mutual in-

2 Multiscale Structure

formation, which is the difference between the total information of two components considered separately and the joint information of those two components taken together:

$$I(a; b) = H(a) + H(b) - H(a, b). \quad (2.2)$$

By extension, we can also define the tertiary mutual information

$$I(a; b; c) = H(a) + H(b) + H(c) - H(a, b) - H(b, c) - H(a, c) + H(a, b, c). \quad (2.3)$$

This can be extended to higher scales in the same fashion, defining *shared information* for sets of four or more components.

One way to understand the meaning of our axioms is the following. Take a set of questions, which are all mutually independent in the sense that answering one doesn't help to answer any other, or any combination of others. Each question pertains to one or more components of a system. Components have nonzero shared information if one or more questions pertain to both of them. In other words, if each question is represented by a point, then each component is a set of points, and set-theoretic intersection defines shared information.

As mentioned earlier, this formalism treats scale and information as complementary quantities. Often, "scale" is thought of in terms of length or time (for example, the James quote above organizes the learned disciplines essentially by the geometrical dimensions of what they study). For an axiomatic development, a more general definition is appropriate, and so for our purposes, "scale" will refer to the *number of components* within a system which are involved in an interrelationship.

2.1.2 Indices of Structure

To specify the structure of a system according to our definition, it is necessary to specify the information content of each subset $U \subset A$. Because the number of such subsets grows exponentially with the number of components, complete descriptions of structure are impractical for large systems. Therefore, we require statistics which can convey the general character of a system's structure without specifying it completely. Using an *index of structure*, we can summarize how a system is organized and compare that pattern of organization to the patterns manifested by other systems.

One such index of structure is the *complexity profile*, introduced in [5] to formalize the intuition that a genuinely complex system exhibits structure at multiple scales. The complexity profile is a real-valued function $C(k)$ that specifies the amount of information contained in interdependencies of scale k and higher. $C(k)$ can be computed using a combinatorial formula which takes as input the values of the information function H on all subsets $U \subset A$ [3, 4, 5, 6, 7, 8]. First, we define the quantity $Q(j)$ as the sum of the joint information of all collections of j components:

$$Q(j) = \sum_{i_1, \dots, i_j} H(a_{i_1}, \dots, a_{i_j}). \quad (2.4)$$

The complexity profile can be computed using the formula

$$C(k) = \sum_{j=N-k}^{N-1} (-1)^{j+k-N} \binom{j}{j+k-N} Q(j+1), \quad (2.5)$$

where $N = |A|$ is the number of components in the system. Generally, $C(k)$ captures the amount of information contained in interrelationships of order k and higher. We shall illustrate this in the next section with a few examples.

The complexity profile satisfies a *conservation law*: the sum of $C(k)$ over all scales k is

$$\sum_k C(k) = \sum_{a \in A} H(a). \quad (2.6)$$

That is, the sum of the complexity over all scales is given by the individual information assigned to each component, regardless of the components' interrelationships.

Another useful index of structure is the Marginal Utility of Information, or MUI [3]. While the complexity profile characterizes the amount of information that is present in the system behavior at different scales, the MUI is based on descriptive utility of limited information through its ability to describe behavior of multiple components. Informally speaking, we describe a system by “investing” a certain amount of information, and for any amount of information invested, an *optimal* description yields the best possible characterization of the system. The MUI expresses how the usefulness of an optimal description increases as we invest more information. We can define the MUI precisely, starting with the basic axioms of information functions, by using notions from linear programming [3].

In general outline, one constructs the MUI as follows. Let $\mathcal{A}$ be a system, defined per our formalism as a set of components A and an information function H . Then, let d be a *descriptor*, an entity which conveys information about the system $\mathcal{A}$. To express this mathematically, we consider the new, larger system made by conjoining d with the set of components A and defining an information function on the subsets of this expanded set. The information function for the augmented system reduces to that of the original for all those interdependencies which do not involve the descriptor d , and it expresses the shared information between d and the original system. The *utility* of d is the sum of the shared information between d and each component within $\mathcal{A}$:

$$u(d) = \sum_{a \in A} I(d; a). \quad (2.7)$$

This counts, in essence, the total scale of the system's organization that is captured by d . We define the *optimal utility* $U(y)$ as the utility of the best possible descriptor having $H(d) = y$. The MUI is then the derivative of $U(y)$.

How do these structure indices capture the organization of a system? We can illustrate the general idea by way of a conceptual example. Consider a crew of movers, who are carrying furniture from a house to a truck. They can be acting largely independently, as when each mover is carrying a single chair, or they can be working in

2 Multiscale Structure

concert, transporting a large item that requires collective effort to move, like a grand piano. In the former case, knowing what any one mover is doing does not say much about what specific act any other mover is engaged with at that time. Information about the crew applies at the scale of an individual mover. By contrast, in the latter case, the behavior of one mover can be inferred from that of another, and information about their actions is applicable at a larger scale. From these general considerations, it follows that for a system of largely independent movers, $C(k)$ is large at small k and drops off rapidly, whereas when the movers are working collectively, $C(k)$ is small for low k and remains nonzero for larger k . When the movers act mostly independently, we cannot do much better at *describing* their behavior than by specifying the behavior of each mover in turn. Therefore, as we invest more information into describing the system, the gain in utility of our description remains essentially constant. For the case of independent movers, then, the MUI curve is low and flat. On the other hand, when the movers are acting in concert, a brief description can have a high utility, so the MUI curve is peaked at the origin and falls off sharply. Heuristically speaking, we can in this example think of the complexity profile and the MUI as *reflections* of each other. When we develop these indices quantitatively, we find in fact that this is exactly true in a broad class of systems.

Both the complexity profile and the MUI obey a convenient sum rule. If a system separates into two independent subsystems, the complexity profile of the whole is the sum of the profiles of the pieces, and likewise for the MUI. This property of both structure indices follows from the basic information-function axioms [3]. In the next section, we will see examples of systems which illustrate the sum rule for both the MUI and the complexity profile.

2.2 Examples

2.2.1 Three-Component Systems

To explore the consequences of our definitions, it is helpful to begin with simple examples. Following the recent review article about the multiscale complexity formalism [3], we study the following four systems, each of which contains three binary variables.

- *Example A: Three independent bits:* The system comprises three components, and knowing the state of any one bit provides no inference about the state of any other. As a whole, the system can be in any one of eight possible configurations, with no preference given to any of the eight possibilities.
- *Example B: Three completely interdependent bits:* The system as a whole is either in state 000 or state 111, with no preference given to either option. Knowing the value of any one bit allows the inference of both other bits.
- *Example C: Independent blocks of dependent bits:* Each component is equally likely to take the value 0 or 1; however, the first two components always take the same value, while the third can take either value independently of the coupled pair.

- **Example D: The 2 + 1 parity bit system:** Three bits which can exist in the states 110, 101, 011, or 000 with equal probability. Each of the three bits is equal to the parity (0 if even; 1 if odd) of the sum of the other two. Any two of the bits are statistically independent of each other, but the three as a whole are constrained to have an even sum.

Figures 2.1 and 2.2 show the complexity profiles and the MUI curves for these example systems.

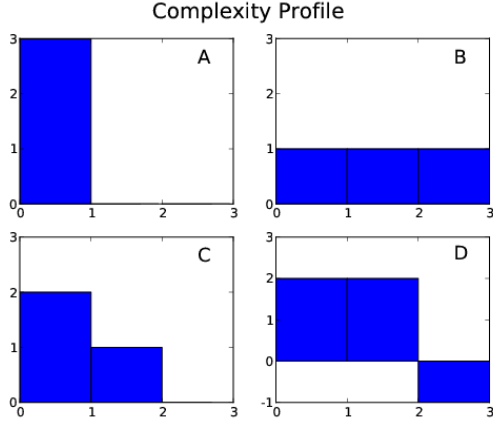


Figure 2.1: Complexity profiles for the three-component example systems **A**, **B**, **C** and **D**, computed using Eq. (2.5). Examples **A** and **B** illustrate the general fact that highly interdependent systems have tall and narrow complexity profiles, whereas the profiles of systems with largely independent components are low and wide. Example **C**, which we can think of as the combination of two independent subsystems, illustrates the complexity profile's sum rule. Finally, example **D**, the parity-bit system, showcases the emergence of negative shared information. Note that the total signed area bounded by each curve is 3 units. (Figure reproduced from [3].)

2.2.2 Minimal Incidence Geometry

To develop additional intuition about our information-theoretic formalism, and to build a bridge between different areas of mathematics, we shall apply the information theory of multicomponent systems to *incidence geometry*. The premise of incidence geometry is that one has a set of points and a set of lines which connect them, satisfying some conditions which abstract basic notions of geometry. To wit, for any incidence geometry, every line contains at least two distinct points, and for every line, there exist one or more points not lying on that line. We relate geometry to information theory in the following way: Ascribe to each point 1 unit of information, and define each line

2 Multiscale Structure

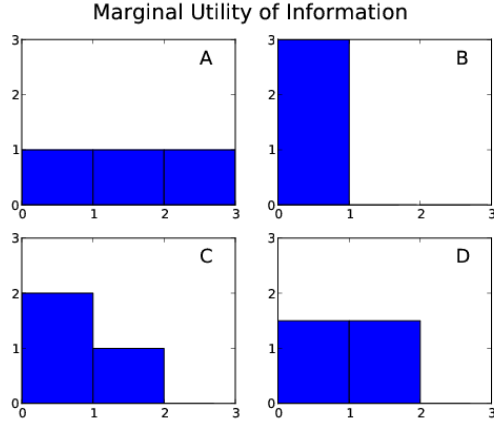


Figure 2.2: MUI plots for the three-component example systems **A**, **B**, **C** and **D**. Note that, as with the complexity profiles in Figure 2.1, the total area bounded by each curve is 3 units. Furthermore, for examples **A**, **B** and **C**, the complexity profile and the MUI are reflections (generalized inverses) of each other. This is generally true for systems which are the disjoint union of internally interdependent blocks [3], but it is not the case for the parity-bit system, example **D**. (Figure reproduced from [3].)

to be a system component. Then for any incidence geometry, the information ascribed to a component is always greater than or equal to 2, and the information within the whole system is always greater than 2.

The examples we shall consider from incidence geometry will illustrate most of the key features of the multiscale information theory formalism. The noteworthy exception is that incidence geometry does not provide examples of negative multivariate mutual information. This is a subtlety which can arise when one considers dependencies among three or more components [3], as we saw in example **D**. However, it will not be a major concern for the models from mathematical biology which we will study later in this chapter.

The simplest possible incidence geometry contains 3 points and 3 lines. We depict this construction in Figure 2.3.

If we denote the three lines by l_1 , l_2 and l_3 , as in Figure 2.3, then because each line contains exactly two points, we have

$$H(l_1) = H(l_2) = H(l_3) = 2, \quad (2.8)$$

while because any two lines intersect in exactly one point,

$$I(l_1; l_2) = I(l_2; l_3) = I(l_1; l_3) = 1. \quad (2.9)$$

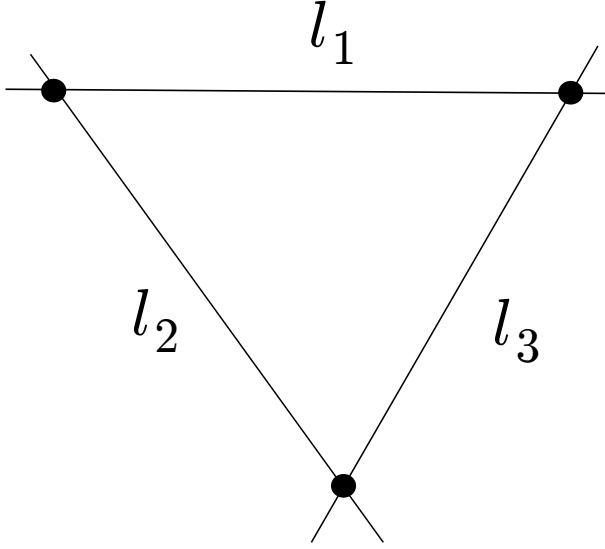


Figure 2.3: The simplest possible incidence geometry. Each of the three lines contains two distinct points, and for each line, there exists exactly one point which does not lie on that line. When we associate a unit of information to each point, the shared information of any pair of lines is the information ascribed to their point of intersection.

The joint information of all three components taken together is the total number of points in the geometry:

$$H(l_1, l_2, l_3) = 3. \quad (2.10)$$

From these three observations, we can deduce that the tertiary mutual information of the three components vanishes:

$$C(3) = I(l_1; l_2; l_3) = 0. \quad (2.11)$$

This is the information-theoretic restatement of the geometric fact that the three lines do not all come together at a single point. All together, the complexity profile of the minimal incidence geometry is given by

$$C(k) = \begin{cases} 3, & k \in \{1, 2\}, \\ 0, & k = 3. \end{cases} \quad (2.12)$$

The information-theoretic relationships among the three system components l_1 , l_2 and l_3 can be expressed in a three-circle Venn diagram, which we depict in Figure 2.4.

The other structure index introduced above is the MUI. We can deduce the MUI curve of the minimal incidence geometry using the properties of the MUI established in [3]. First, a descriptor d must have at least 3 units of information to capture all

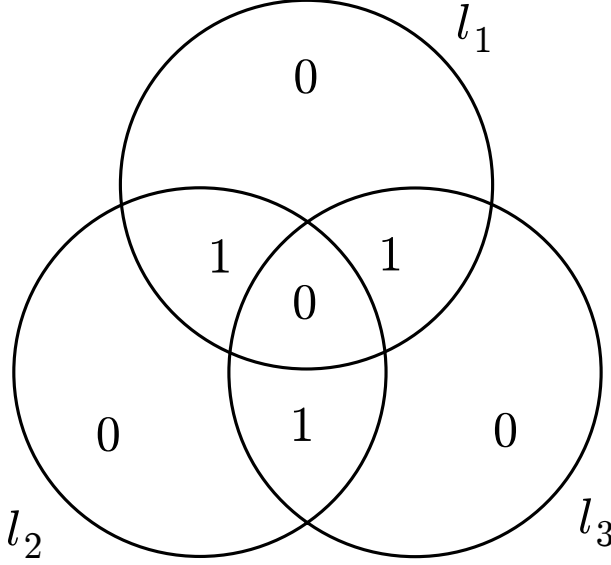


Figure 2.4: Information diagram for the minimal incidence geometry depicted in Figure 2.3. Within each of the three circles, all of the regions which contain nonzero information are regions of overlap with another circle. This is the information-theoretic consequence of the geometrical fact that each point belongs to more than one line. Note that the central region, where all three circles overlap, contains no information.

of the information which was granted to the geometry. Expanding on an *optimal* descriptor, one which wastes nothing, brings no benefit beyond a descriptor length of $H(d) = 3$. Therefore, the marginal utility $M(y)$ will equal zero for $y \geq 3$. In addition, because the integral of the MUI curve is the utility of a full description—that is, the total scale-weighted information of the system—we know the integral of the MUI for this geometry will be 6. Furthermore, we can constrain the height of the MUI curve, using the following property:

- If there are no interactions or correlations of degree k or higher—formally, if $I(a_1; \dots; a_k) = 0$ for all collections $a_1, \dots, a_k$ of k distinct components—then $M(y) \leq k$ for all y [3, §VII.B].

Here, this means that $M(y) \leq 3$. Furthermore, we know that $M(y)$ is the derivative of a piecewise linear function, so $M(y)$ is piecewise constant. For the minimal geometry, then, we expect the MUI should be

$$M(y) = \begin{cases} 2, & 0 \leq y < 3, \\ 0, & y \geq 3. \end{cases} \quad (2.13)$$

Note that the MUI curve is the reflection of the complexity profile $C(k)$ in Eq. (2.12).

These properties hold for any incidence geometry: the MUI vanishes for y larger than the number of points used to build the geometry, the integral of the MUI is the total scale-weighted information, and the MUI is bounded above by one plus the maximal number of lines which mutually intersect at a common point. We can relate the MUI and the complexity profile by noting that the areas bounded by both curves are the same, and moreover, the width of the MUI curve $M(y)$ is the height of the complexity profile $C(1)$, because both are given by the number of points in the geometry.

2.2.3 Fano Plane

The Fano plane, pictured in Figure 2.5, has 7 points, 7 lines, 3 points on every line, and 3 lines through every point. The total scale-weighted information is the number of points per line times the number of lines, or 21. The information content of the whole system is 7, while the mutual information between any line and any other is again 1.

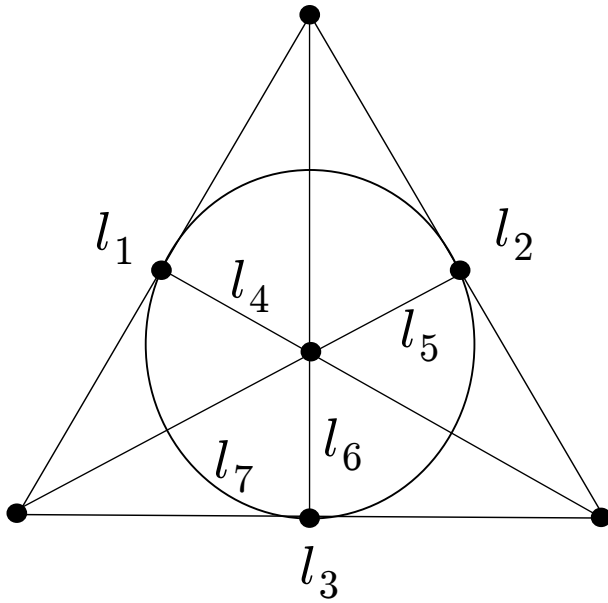


Figure 2.5: The Fano plane: a symmetrical arrangement of seven points and seven lines. The shared information between any two lines is 1, but the shared information at higher scales depends on which set of lines we choose.

For the Fano plane, there are two possible information-theoretic scenarios involving three distinct lines. If the three lines do not all meet at a common point, as for example l_1 , l_2 and l_3 , then the tertiary mutual information of those three components is zero. The other option is for the three lines to meet at a common point, as with l_1 , l_3 and

2 Multiscale Structure

l_5 , in which case the tertiary mutual information is 1.

$$I(l_1; l_2; l_3) = 0, \text{ but } I(l_1; l_3; l_5) = 1. \quad (2.14)$$

The tertiary mutual information $I(l_i; l_j; l_k)$ can never be greater than 1, because any two lines come together in one and exactly one point. This uniqueness of intersections also implies that, in the three-circle Venn diagram, the lens-shaped regions where two circles overlap always contain a total of 1 unit of information. If the inner region of triple overlap, corresponding to $I(l_i; l_j; l_k)$, contains the value 1, then the outer region, $I(l_i; l_j | l_k)$, must contain the value 0, and vice versa. In addition, the total information content enclosed by each of the three circles is always 3 units, the number of points per line. Together, these facts constrain the possible three-circle Venn diagrams for subsets of the Fano-plane system, leaving only the two possibilities depicted in Figure 2.6.

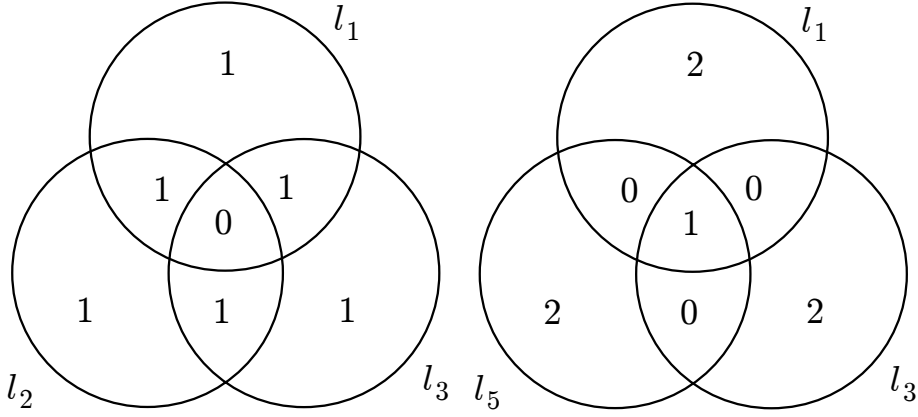


Figure 2.6: Illustrative examples of the two possible three-circle Venn diagrams for three-component subsets of the Fano-plane system. Note that the information in the central region is zero in one case but nonzero in the other.

We have in the Fano plane an elementary illustration of a commonplace occurrence in complex systems: *higher-order structure which cannot be resolved into lower-order interrelationships*. In this case, we see there exists a variety among triples of components which cannot be inferred from considering pairs. The Fano-plane system does not fall within the particular special classes of systems studied as examples in [3], since not all subsets of the same size have the same information content.

We can deduce the complexity profile $C(k)$ of the Fano-plane system by counting points lying on k or more lines, or by computation using Eq. (2.5). If we take the latter approach, with 7 components there are $2^7 - 1 = 127$ nonempty subsets U for which we must find the information content $H(U)$. However, thanks to the high degree of symmetry, the information content for the different possible subsets of the component set is not that hard to work out. Because there exist 3 lines through any point, eliminating 1 line (component) can only reduce the number of lines through any

2.3 The Dual of a Complex System

point down to 2. So, the information content of any six-component set is still 7. To eliminate all lines through a point, we must erase at least three components. From considerations like these, we can deduce $C(k)$ by explicit computation with Eq. (2.5). For the Fano plane, the complexity profile is given by

$$C(k) = \begin{cases} 7, & 1 \leq k \leq 3, \\ 0, & 4 \leq k \leq 7. \end{cases} \quad (2.15)$$

We can also deduce this result quickly by recalling that, generally, $C(k)$ captures the amount of information contained in interrelationships of scale k and higher. In the context of incidence geometry, the complexity profile $C(k)$ has the geometrical interpretation as *the number of distinct points which lie at the intersection of k or more lines*.

Applying the same properties as we used for the minimal incidence geometry, we can deduce that the MUI of the Fano plane vanishes for descriptor lengths $y \geq 7$, that the integral of $M(y)$ is $7 \cdot 3 = 21$, and that $M(y) \leq 4$ for all y .

2.3 The Dual of a Complex System

Geometry makes much use of *duality*. The dual of a geometrical arrangement is that arrangement which is found by interchanging the roles of points and lines in the original. For example, an *affine plane of order n* is a specialization of an incidence geometry in which the following conditions hold [25, 26]:

- the geometry contains n^2 points in all;
- the geometry contains $n(n + 1)$ lines;
- each line contains n points;
- each point lies on $n + 1$ lines.

Interchanging points and lines yields a *dual affine plane* [25, 27], a geometry which meets the following criteria:

- the geometry contains $n(n + 1)$ points;
- the geometry contains n^2 lines;
- each line contains $n + 1$ points;
- each point lies on n lines.

If we translate from geometry to information theory, what is the meaning of duality? We began by saying that each point in a geometry corresponded to a unit of information, and each line was to become a component in an information-theoretic system. Applying the operation of duality, we find that each line in the original geometry

2 Multiscale Structure

should be ascribed a unit of information in the dual system, and each point in the original geometry becomes a system component in the dual.

The complexity profile of the original system is defined for values of k from 1 to the number of lines in the original geometry. Therefore, the complexity profile of the dual system is defined for values of k from 1 to the number of *points* in the original geometry. The property of “lines meeting at a common point” in the original becomes the property of “points lying on a common line” in the dual. Consequently, the complexity profile of the dual system can be found from the original geometry. The dual of k lines intersecting at a point is k points sharing a common line. We can find the complexity profile of the dual system by counting the number of distinct lines in the original system which contain k or more points.

For the minimal incidence geometry depicted in Figure 2.3 and the Fano plane portrayed in Figure 2.5, the duality exchange operation does not change the complexity profile. We can say that $C(k)$ for those geometries is *self-dual*. Considering the affine planes of order n defined above, we know that each point lies on $n + 1$ lines, so $C(k)$ is a rectangle of width $n + 1$ and height n^2 . In a dual affine plane of order n , each of the $n(n + 1)$ points lie on n lines, so $C(k)$ is a rectangle of width n and height $n(n + 1)$. For affine planes, the duality transformation preserves the *area*, but not the *shape*, of the complexity profile.

One naturally wonders whether this property of the area under the $C(k)$ curve is more general. We can investigate this question using the conservation law, proved in Allen *et al.* [3], that the area under the complexity profile is always the total scale-weighted information content of the system. For the special case of an incidence geometry, this means that

$$\sum_k C(k) = \sum_i H(l_i), \quad (2.16)$$

where i ranges from 1 to the total number of lines in the geometry. Let $\{v_j\}$ be the set of all points in the geometry. (For an affine plane, j thus ranges from 1 to n^2 .) The area bounded by the complexity profile is then

$$\sum_i H(l_i) = \sum_i \sum_{\{j|v_j \in l_i\}} H(v_j). \quad (2.17)$$

In this sum, the multiplicity of any $H(v_j)$ is the number of lines which contain the point v_j . Therefore,

$$\sum_i H(l_i) = \sum_j H(v_j) \cdot |\{l_i | v_j \in l_i\}|. \quad (2.18)$$

The components of the dual system are the points of the original geometry, and the information content of each component in the dual system is the number of lines which pass through that point in the original geometry. Summing over the components of the dual system to find the total area under the dual complexity profile yields the same sum as in Eq. (2.18). In consequence, we can say that *the duality transformation generally preserves the area under the complexity profile*.

2.3 The Dual of a Complex System

What can we deduce about the Marginal Utility of Information for affine planes and their duals? We recall that the MUI is, by construction, a piecewise constant function. The general properties we deduced earlier imply that for an affine plane of order n , $M(y)$ vanishes for $y \geq n^2$, and the integral of $M(y)$ is the number of lines times the number of lines per point, or $n^2(n+1)$. Furthermore, because exactly $n+1$ lines meet at each point, $M(y) \leq n+2$ for all y .

The most straightforward way to meet these requirements is with the following piecewise constant curve:

$$M(y) = \begin{cases} n+1, & 0 \leq y < n^2, \\ 0, & y \geq n^2. \end{cases} \quad (2.19)$$

For a *dual* affine plane of order n , the integral of $M(y)$ is the same, $n^2(n+1)$. The right-hand edge is instead at $n(n+1)$, and the upper limit becomes $M(y) \leq n+1$. So,

$$M(y) = \begin{cases} n, & 0 \leq y < n(n+1), \\ 0, & y \geq n(n+1). \end{cases} \quad (2.20)$$

For both affine planes and their duals, $M(y)$ is the reflection of the complexity profile $C(k)$. And again, the duality transformation preserves areas but not shapes.

These self-duality properties are the information-theoretic consequences of the geometry theorem known as the *principle of double counting* [26]. To wit: if $\{v_i\}$ are the points of an incidence geometry and $\{l_j\}$ are its lines, then

$$\sum_{\{v_i\}} |\{l_j \mid v_i \text{ lies on } l_j\}| = \sum_{\{l_j\}} |\{v_i \mid l_j \text{ intersects } v_i\}|. \quad (2.21)$$

We can prove this by observing that both sides of Eq. (2.21) are equal to the size of the set of ordered pairs defined by

$$S = \{(v_i, l_j) \mid v_i \text{ and } l_j \text{ are incident}\}. \quad (2.22)$$

2.4 Computation and Gammoids

The idea of information is closely related to that of *computation*, and in this section, we will explore another type of system for which we can define an information function, thereby bringing our concepts of multiscale structure into a computational context.

We can idealize a computation as a mathematical process that takes a set of *inputs* to a set of *outputs*. For example, if we have a program that takes a number and returns its square, we can represent that code pictorially as an arrow:



We have drawn the output as a filled circle, and the input as an unfilled one. The picture (2.23) is a diagrammatic representation for any function that acts on a single input to produce one and only one output: the sine of an angle, the number of meters in a distance specified in furlongs and so on.

Some functions take two inputs and return a single output. Given two numerical variables, for example, we could compute their sum, their product, the logarithm of one to the base given by the other, and so forth. We can represent all of these functions pictorially with the following diagram:



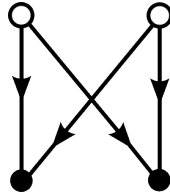
Alternatively, a single input might be used to create multiple outputs. For example, we might be given the time elapsed in seconds since midnight, and calculate the current hour and the minute within that hour:



There is a sense in which these outputs are not independent, for they derive from

a common source. This is true even if, as is the case for separating clock time into minutes and hours, we cannot infer one output from the other.

What other patterns of information flow might we encounter as we go about our business? Given two positive integers, we can find their greatest common divisor and their least common multiple. If a function performs both of these tasks, then it has two outputs, both of which depend on the two inputs:



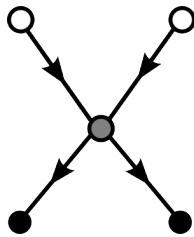
(2.26)

A computation may proceed by way of intermediate steps:



(2.27)

These intermediate stages may combine multiple inputs, and they can yield multiple outputs. In our diagrams, input vertices (unfilled circles) have no incoming links, and output vertices (filled circles) have no outgoing ones. We introduce intermediate vertices with both types of links, as in the following:



(2.28)

In this example, two pieces of input data are combined to yield an intermediate result, which is then used to compute two output values. If we saved this intermediate result, we could reconstruct the two outputs without having the original input data. It might be that we cannot deduce the left-hand output from the right, but there is still an interdependency between them, linking them by virtue of their common past.

2 Multiscale Structure

Consider a computation with two inputs and two outputs, one of which is arrived at by means of an intermediate step:



Both of the outputs derive ultimately from the same inputs. However, they are in a way independent of each other, because knowing the result of the intermediate step lets us compute one output but not the other. Diagrams with multiple layers of intermediate steps are also possible, and they have a straightforward interpretation in terms of computations that proceed in stages.

These considerations motivate the following idea: The *effective amount of information* associated with a set of outputs is the number of inputs and/or intermediate values which one must know in order to produce those outputs. We are, after a fashion, giving a mathematical form to the notion that the difficulty of a computational task is how arduous it would be to recover from a crash!

All of our diagrams have taken the form of *directed graphs*. Each one is a set of vertices, connected by a set of edges, where each edge carries an indication of which vertex is its source and which is its target. In addition, we have distinguished subsets of the vertex set, marking each vertex as input, intermediate or output. Recognizing this, we can develop the idea more generally.

Let G be a directed graph, or “digraph” for short. Designate a subset S of its vertices as inputs, and select a subset T of vertices to be the outputs. We can think of S as the set of “sources” for information flow, and T as the set of its “targets.” (In the digraphs we have drawn, the vertices in S have no incoming edges, and those vertices in T have no outgoing edges. This is sensible, but it turns out not to be essential for proving the key properties of the information function.) The information content $H(U)$ of a subset $U \subset T$ is the size of the smallest set of vertices having the property that all paths from S to U must pass through it. The size of this “minimal separating set” tells us how many intermediate variables we would need to save in order to compute the outputs in U .

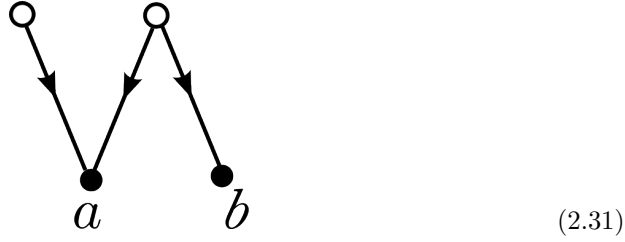
It can be proven that the function H satisfies our axioms for being an information function. Consequently, the function H yields sensible expressions for the interdependence of output variables. Having defined H , we can construct the indices of multiscale structure $C(k)$ and $M(y)$ as we did before.

For example, take the graph



Here, we have one output, which we label a , and we see directly that $H(a) = 2$.

A graph with two outputs might look like the following:



In this case, $H(a) = 2$, while $H(b) = 1$ and $H(a, b) = 2$. The shared information between outputs a and b is

$$I(a; b) = H(a) + H(b) - H(a, b) = 1, \quad (2.32)$$

corresponding nicely to the fact that a and b depend upon one common input.

A more involved two-output graph could include an intermediate step:



Here, $H(a) = H(b) = 1$, while $H(a, b) = 2$. The shared information between a and b is now $I(a; b) = 0$, because the intermediate stage “shields” output a from the original input.

What we have constructed here is known in pure mathematics as a *gammoid*, and our information function is the *rank function* of that gammoid [28]. A gammoid is an example of a *matroid*, a structure defined as a set of elements M and a rank function that assigns a nonnegative integer to each subset of M . Matroid rank functions are monotonic and strongly subadditive, so they count as information functions. As we have seen in this section, matroid theory furnishes examples of mathematical entities

to which we can apply the ideas of multiscale structure.

2.5 Network Dynamics

We can apply the multiscale complexity formalism quantitatively to a model which is an idealized representation of multiple interesting biological scenarios. Doing so requires relating the concepts of probability and information; we provide a more detailed development of the mathematical prerequisites in Chapter 5.

One way to make progress on many biological problems is to make the approximation that each component of a system can, at any given time, be in one of two mutually exclusive states of being. In essence, we idealize a phenomenon by treating it as composed of binary random variables, possibly correlated. We might postulate that each organism in a population can follow one of two survival strategies. For example, a male bower bird can maraud, attacking other birds' bowers, or it can remain at its own bower, guarding its own mating display from marauders [29, 30]. Or, we might postulate that a gene comes in two variant forms. Each instance of the gene in the idealized population is then a binary random variable. We can make an analogous approximation when modeling social and economic systems. For example, an individual voter can choose between one of two political parties. Or, in a simplified but still instructive model for a stock market, the price of a company's stock can be going either up or down [31].

A specific implementation of this idea is the *Moran model*, which was originally formulated in biology but can be applied more broadly [31]. Consider a haploid population of N individuals, and a gene which comes in two alleles. The genetic character of the population can change as individuals are born and die. One simple dynamical model for this process picks an individual at random with each tick of a discrete-time clock. The chosen, or *focal*, individual mates with one of the other $N - 1$ organisms and produces an offspring, which then takes the place of the focal individual. The allele carried by the offspring is that carried by one of its two parents, the choice of which parent being made randomly with equal probability either way.

Reframing the Moran model in network-theory language turns out to be convenient for developing extensions, such as treatments including structured populations, wherein mating is not uniformly random. Furthermore, doing so broadens the range of systems to which the mathematics can be applied: moving away from the specifically biological terminology makes it more explicit that the Moran model can be applied equally well to biological evolution or to social dynamics [31].

The components of our system will be the N nodes of a network. Each node is a random variable which can take the values 0 and 1. In addition to these N nodes, we augment the system with a number N_0 of nodes whose states are all fixed at 0, and a quantity N_1 of nodes whose states are fixed to be 1.

At each time step, we pick one of the variable nodes at random. We then choose, stochastically, whether or not to change that node's value. With probability p , we keep the node value the same, and with probability $1 - p$, we assign to it the value of another node, chosen at random from a pool of candidates. This pool contains

both the neighborhood of the node in the network topology and the $N_0 + N_1$ fixed nodes. In this way, the fixed nodes represent the possibility of mutation: even if all the dynamical population has allele 1, there remains the opportunity of picking up a 0, and vice versa. For a complete graph, the steady-state behavior of this dynamical system can actually be found analytically [32]. The probability that exactly m nodes out of the N whose value can vary will be in state 1 is

$$q(m) = A(N, N_0, N_1) \frac{\Gamma(N_1 + m) \Gamma(N + N_0 - m)}{\Gamma(N - m + 1) \Gamma(m + 1)}, \quad (2.34)$$

where the normalization constant A is given by

$$A(N, N_0, N_1) = \frac{\Gamma(N + 1) \Gamma(N_0 + N_1)}{\Gamma(N + N_0 + N_1) \Gamma(N_0) \Gamma(N_1)}. \quad (2.35)$$

We illustrate $q(m)$ for networks of $N = 10$ nodes and different values of N_0 and N_1 in Figure 2.7. The function $q(m)$ is an example of the *beta-binomial distribution*, which is significant in probability theory for reasons we will return to in Chapter 5.

Because the gamma function can take noninteger values, we can compute the probability $q(m)$ even for nonintegral N_0 and N_1 . This is useful if we wish to examine the low-mutation-rate limit.

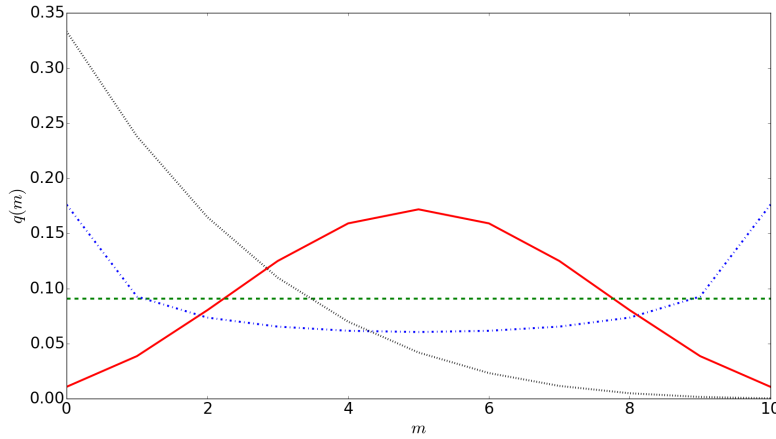


Figure 2.7: Probability that exactly m nodes out of 10 will be in the 1-state, for different values of N_0 and N_1 . Red (solid): $N_0 = N_1 = 5$; green (dashed): $N_0 = N_1 = 1$; blue (dash-dotted): $N_0 = N_1 = 0.5$; black (dotted): $N_0 = 5$ and $N_1 = 1$.

If the network topology is that of a complete graph, then the system has *exchange symmetry*, an invariance under permutations which simplifies the calculation of struc-

2 Multiscale Structure

ture indices [3]. This simplification follows from the fact that if exchange symmetry holds, all subsets having the same number of components can be taken to contain the same quantity of information. Formally, for each set $U \subset A$, the information of U is a function of the cardinality $|U|$, which we can write as a subscript, $H(U) = H_{|U|}$.

Recalling that the complexity profile $C(k)$ indicates the information in dependencies of scale k and higher, the information specific to scale k is

$$D(k) = C(k) - C(k+1). \quad (2.36)$$

The sum of $D(k)$ over all scales k is $C(1)$. For any fixed scale k , the complexity $D(k)$ is (up to a prefactor) the binomial transform of the sequence $a_l \equiv H_{l+N-k}$.

$$D(k) = \binom{N}{k} \sum_{l=0}^k (-1)^{l+1} \binom{k}{l} H_{l+N-k}. \quad (2.37)$$

We can calculate H_n from the probability distribution $q(m)$, as given by Eq. (2.34). Knowing H_n , we can compute $D(k)$, from which we can easily find the complexity profile $C(k)$. Figure 2.8 illustrates the results. We see that $C(k)$ depends upon the numbers of fixed influence nodes, N_0 and N_1 . When $C(k)$ is concentrated at $k = 1$, the nodes are changing their values almost independently of one another. This is the case, for example, when we set $N_0 = N_1 = 5$. For those parameter values, the external influences are stronger than those of the variable nodes upon each other, while being equally balanced in both directions. This creates a situation in which knowing the status of any one variable node provides very little information about the status of any other. On the other hand, when $C(k)$ is elevated at larger values of k , then nonnegligible amounts of information apply at higher scales. This occurs when the external influences are weaker than the internal dynamics, causing the variable nodes to act collectively.

2.6 Frequency-Dependent Moran Process

One fundamental fact of evolutionary biology is that the environment of an organism consists in large part of other organisms. A simple, albeit approximate, way to represent the configuration of an ecosystem is by specifying the frequencies of abundance or population densities for the species which are present. (We will consider more sophisticated approximations, and the hazards of oversimplified representations, in later chapters.) In this context, we can speak of *frequency-dependent fitness*: the success of an organism type or an evolutionary strategy can be a function of the current population densities.

The simplest kind of frequency dependence is a linear relationship between population density and fitness. As before, we consider two varieties, and we keep the total population size constant, so the frequencies of both types can be given in terms of a single variable x . We take the reproductive rates of type-0 and type-1 organisms to

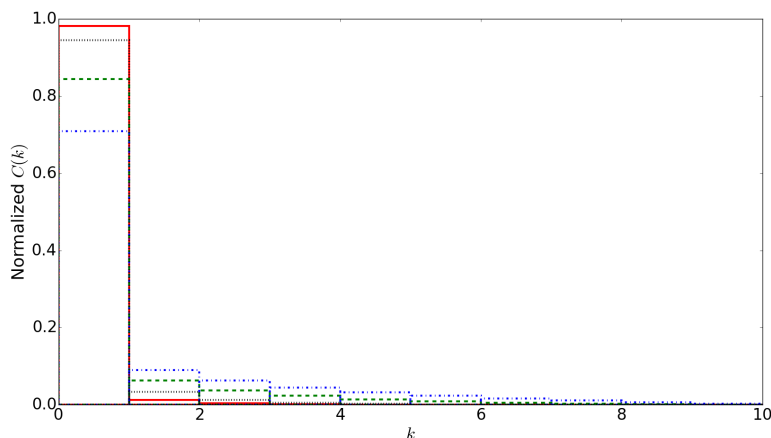


Figure 2.8: Complexity profiles for the cases illustrated in Figure 2.7. Each curve is normalized so that the total area under it is 1. Elevated complexity C at larger k indicates collective behavior at larger scales. Red (solid): $N_0 = N_1 = 5$; green (dashed): $N_0 = N_1 = 1$; blue (dash-dotted): $N_0 = N_1 = 0.5$; black (dotted): $N_0 = 5$ and $N_1 = 1$.

be given by

$$f_0(x) = A_{01}x + A_{00}(1 - x), \quad (2.38)$$

$$f_1(x) = A_{11}x + A_{10}(1 - x). \quad (2.39)$$

The coefficient A_{ij} is the payoff which a type- i player gains by playing with a type- j player. Different values of the matrix A represent different interactions between evolutionary strategies.

In order to apply Shannon information theory, we need a probability distribution. A convenient and meaningful one for these purposes is the *mutation-selection equilibrium* distribution, which is the steady state of the frequency-dependent Moran process. We can find this distribution numerically by iterating the appropriate update rule, which we can represent as multiplication by a transition matrix. The next step is to construct this matrix. Having done so, we will be able to compute $q(m)$ and thence obtain the complexity profile, as before. The result will typically depend both upon the payoff matrix A and on the mutation rate.

Let the total population size be N , and let m denote the number of type-1 individuals. We suppose that reproduction is imperfect, with mutations occurring at rate u . That is, an offspring inherits its parent's type with probability $1 - u$, while with probability u , we pick the offspring's type at random. A nonzero mutation rate implies that the population does not have to get stuck in a uniform configuration: even if all individuals have the same type, an error in reproduction can create an organism of

2 Multiscale Structure

the opposite type in the next generation. This is a necessary requirement for having a steady-state probability distribution which is not concentrated entirely at $m = 0$ or $m = N$.

To find the steady-state probability distribution for m , we first need to calculate the probabilities that m will increase or decrease. In the frequency-dependent Moran process [33], the probability that m will decrease by 1 is

$$p_{m \rightarrow m-1} = \frac{m}{N} \left((1-u) \frac{(N-m)f_0\left(\frac{m}{N}\right)}{mf_1\left(\frac{m}{N}\right) + (N-m)f_0\left(\frac{m}{N}\right)} + \frac{u}{2} \right). \quad (2.40)$$

And the probability of m increasing by 1 is

$$p_{m \rightarrow m+1} = \frac{N-m}{N} \left((1-u) \frac{mf_1\left(\frac{m}{N}\right)}{mf_1\left(\frac{m}{N}\right) + (N-m)f_0\left(\frac{m}{N}\right)} + \frac{u}{2} \right). \quad (2.41)$$

With these equations, we can find the steady-state probability distribution $q(m)$, which will depend on the payoff matrix A and the mutation rate u . (For the present purposes, a numerical computation will suffice.) Knowing $q(m)$, we can as before find the complexity profile $C(k)$. The resulting curve tells us about the scales of organization which arise within the population as a consequence of the evolutionary game dynamics.

To connect with the literature [33], we carry out this calculation for the payoff matrix

$$A = \begin{pmatrix} 6 & 4 \\ 7 & 5 \end{pmatrix}, \quad (2.42)$$

which defines an instance of the *Prisoner's Dilemma*. One application of this to biology is the case of the bower birds mentioned earlier [29, 30]. Simplifying somewhat, a male bower bird has two strategies available to it: to guard its own bower, or to maraud and attempt to damage others. Designate guarding as strategy 0 and marauding as strategy 1. The matrix element A_{ij} denotes the payoff to a bird following strategy i against an opponent who plays strategy j . In this example, a guardian (row 0) who plays against a marauder (column 1) obtains a score of 4. The highest payoff is A_{10} , the score obtained by a marauder who plays against a guardian. In fact, it is better to maraud than to guard, when facing either kind of foe:

$$A_{10} > A_{00}, \text{ and also } A_{11} > A_{01}. \quad (2.43)$$

So far, it looks like the thing to do is to maraud. However, the payoff obtained when both birds follow this strategy is A_{11} , which is *less* than the payoff A_{00} they would have obtained if they had both stayed home.

The particular choice of numbers here is arbitrary, but the relationships between the numbers are representative of typical conditions in the wild. As Gonick [30] summarizes, “Seemingly forced by the game’s logic into a hostile strategy, they end up worse off than if they had only cooperated!” A wide variety of biological scenarios can be considered as examples of this game [34, 35]. A primary concern is to identify the

conditions under which cooperation (for example, both bower birds guarding rather than marauding) is evolutionarily favorable.

This type of situation is designated a “Prisoner’s Dilemma” because it is usually introduced with an example of two people apprehended for a crime and interrogated by the police. Each player can choose to say nothing, or to inform on the other player. The payoff matrix is such that it is better to inform than to stay silent, whatever option the other player takes; however, if both players keep quiet, they fare better than if they both inform on each other.

Figure 2.9 shows the probability distribution for the Moran process in mutation-selection equilibrium with this payoff matrix, given two different mutation rates. Note that the effect of varying the mutation rate is quite dramatic. As before, we can compute the complexity profile, which we plot in Figure 2.10.

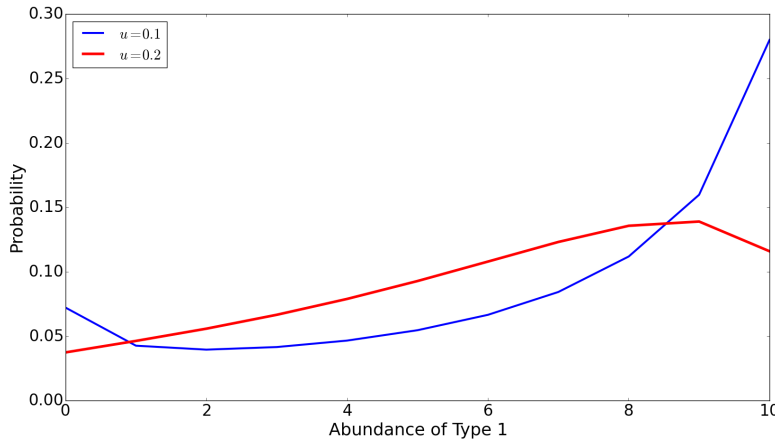


Figure 2.9: Equilibrium probability that exactly m agents out of 10 will be in the 1-state (marauding), for the Prisoner’s Dilemma game defined by Eq. (2.42). Blue (thinner line): $u = 0.1$; red (thicker line): $u = 0.2$.

2.7 Multiscale Challenges and Evolution

In the previous section, we considered the scales of organization which can arise as an evolutionary process develops stochastically. We can also apply our mathematical formalism of multiscale structure to other aspects of evolutionary theory. The following is excerpted from an article by Allen, Bar-Yam and myself [3].

The discipline of cybernetics, an ancestor to modern control theory, used Shannon’s information theory to quantify the difficulty of performing tasks, a topic of relevance both to organismal survival in biology and to system

2 Multiscale Structure

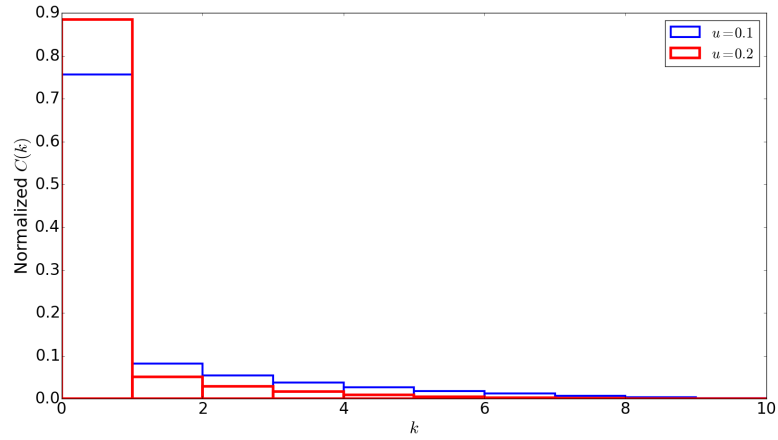


Figure 2.10: Complexity profiles for the cases illustrated in Figure 2.9. Each curve is normalized so that the total area under it is 1. Elevated complexity C at larger k indicates collective behavior at larger scales. Blue (thinner line): $u = 0.1$; red (thicker line): $u = 0.2$.

regulation in engineering. Cyberneticist W. Ross Ashby considered scenarios in which a regulator device must protect some important entity from the outside environment and its disruptive influences [36]. In Ashby’s examples, each state of the environment must be matched by a state of the regulatory system in order for it to be able to counter the environment’s influence on a protected component. Successful regulation implies that if one knows only the state of the protected component, one cannot deduce the environmental influences; *i.e.*, the job of the regulator is to minimize mutual information between the protected component and the environment. This is an information-theoretic statement of the idea of homeostasis. Ashby’s “Law of Requisite Variety” states that the regulator’s effectiveness is limited by its own information content, or *variety* in cybernetic terminology. An insufficiently flexible regulator will not be able to cope with the environmental variability. A multiscale extension of Shannon information theory provides a multiscale cybernetics, with which we can study the scenarios in which “that which we wish to protect” and “that which we must guard against” are each systems of many components, as are the tools we employ for regulation and control [4, 5, 6].

Multiscale information theory enables us to overcome a key limitation of the requisite variety concept. In the examples of traditional cybernetics [36], each action of the environment requires a specific, unique reaction on the part of the regulator. This neglects the fact that the impact which an event in the environment has on the system depends upon the *scale*

2.7 Multiscale Challenges and Evolution

of the environmental degrees of freedom involved. There is a great difference between large-scale and fine-scale impacts. Systems can deflect fine-scale impacts without needing to specifically respond to them, while they need to respond to large-scale ones or perish. For example, a human being can be indifferent to the impact of a falling raindrop, whereas the impact of a falling rock is much more difficult to neglect, even if specifying the state of the raindrop and the state of the rock require the same amount of information. An extreme case is the impact of a molecule: air molecules are continually colliding with us, yet the only effects we have to cope with actively are the large-scale, collective behaviors like high-speed winds. Ashby's Law does not make this distinction. Indeed, there is no framework for the discussion due to the absence of a concept of scale in the information theory he used: Each state is equally different from every other state and actions must be made differently for each different environment.

Thus, in order to account for the real-world conditions, a multiscale generalization of Ashby's Law is needed. According to such a Law, the responses of the system must occur at a scale appropriate to the environmental change, with larger-scale environmental changes being met by larger-scale responses. As with the case of raindrops colliding with a surface, large-scale structures of a system can avoid responding dynamically to small-scale environmental changes which cause only small-scale fluctuations in the system.

Given a need to respond to larger-scale changes of the environment, coarser-scale descriptions of that environment may suffice. A regulator that can marshal a *large-scale response* can use a coarse-grained description of the environment to counteract large-scale fluctuations in the external conditions. In this way, limited amounts of information can still be useful. To make requisite variety a practical principle, one must recognize that information applies to specific scales.

Ashby aimed to apply the requisite variety concept to biological systems, as well as technological ones. An organism which lacks the flexibility to cope with variations in its environment dies. Thus, a mismatch in variety/complexity is costly in the struggle for survival, and so we expect that natural selection will lead to organisms whose complexity matches that of their environment. However, "the environment" of a living being includes other organisms, both of the same species and of others. Organisms can act and react in concert with their conspecifics, and the effect of any action taken can depend on what other organisms are doing at the same time [37]. In some species, such as social insects [38], distinct scales of the individual, colony and species are key features characterizing collective action. This suggests a multiscale cybernetics approach to the evolution of social behavior: We expect that scales of organization within a population—the scales, for example, of groups or colonies—will evolve to match the scales of the challenges which the environment presents. Furthermore, the concept of multiscale response applies within the individual organism as well.

2 Multiscale Structure

Multiple scales of environmental challenges are met by different scales of system responses. To protect against infection, for example, organisms have physical barriers (*e.g.*, skin), generic physiological responses (*e.g.*, clotting, inflammation) and highly specific adaptive immune responses, involving interactions among many cell types, evolved to identify pathogens at the molecular level. The evolution of immune systems is the evolution of separate large- and small-scale countermeasures to threats, enabled by biological mechanisms for information transmission and preservation [39]. As another example, the muscular system includes both large and small muscles, comprising different numbers of cells, corresponding to different scales of environmental challenge (*e.g.*, pursuing prey and escaping from predators versus chewing food) [40].

In Chapter 4, we will use evolutionary game theory to understand one example of multiscale requisite variety: a scenario in which reproductive success depends on multiple organisms acting in concert.

3 Host–Consumer Evolution by Simulation

3.1 Introduction

Spatial extent is a complicating factor in mathematical biology. The possibility that an action at point A cannot immediately affect what happens at point B creates the opportunity for spatial nonuniformity. This nonuniformity must change our understanding of evolutionary dynamics, as the same organism in different places can have different expected evolutionary outcomes. Since organism origins and fates are both determined locally, we must consider heterogeneity explicitly to determine its effects. We use simulations of spatially extended host–pathogen and predator–prey ecosystems to reveal the limitations of standard mathematical treatments of spatial heterogeneity. Our model ecosystem generates heterogeneity dynamically; an adaptive network of hosts on which pathogens are transmitted arises as an emergent phenomenon. The structure and dynamics of this network differ in significant ways from those of related models studied in the adaptive-network field. We use a new technique, organism swapping, to test the efficacy of both simple approximations and more elaborate moment-closure methods, and a new measure to reveal the timescale dependence of invasive-strain behavior. Our results demonstrate the failure not only of the most straightforward (“mean field”) approximation, which smooths over heterogeneity entirely, but also of the standard correction (“pair approximation”) to the mean field treatment. In spatial contexts, invasive pathogen varieties can prosper initially but perish in the medium term, implying that the concepts of reproductive fitness and the Evolutionary Stable Strategy have to be modified for such systems.

Mathematical modeling of biological systems involves a tradeoff between detail and tractability. Here, we consider evolutionary ecological systems with spatial extent—a complicating factor. Analytical treatments of spatial systems typically treat as equivalent all configurations with the same overall population density, the same allele frequencies, the same pairwise contact probabilities or the like. For ease of analysis, one seeks a simplified analytical model, which *coarse-grains* “microstates” (the complete specification of each organism) to “macrostates” (characterized by quantities like average densities), allowing one to make useful predictions about the model’s behavior [41, 42]. Corrections to simple coarse-grainings can quickly generate an overbearing quantity of algebra. It is fairly well appreciated that the simplest approximations break down in the spatial context. What is less acknowledged and not yet systematically understood is that the extensions of the simpler approximations also fail. Before exhausting ourselves with ever-more-elaborate refinements, it would be useful to have

3 Host–Consumer Evolution by Simulation

some understanding of when a particular series of approximations is doomed to inadequacy.

In this chapter, we study the context in which commonly-used coarse-grainings can be expected to fail at capturing the evolutionary dynamics of an ecosystem, and in addition we provide a novel, direct demonstration of that failure. The fundamental issue is *spatial heterogeneity*, a long-recognized concern for mathematical biology [43, 44]. When does spatial heterogeneity significantly impact the choice of appropriate mathematical treatment, and when does a chosen mathematical formalism not capture the full implications of spatial variability? We show that one can test a treatment of heterogeneity by transplanting organisms within a simulated ecosystem in such a way that, were the treatment valid, the modeled behavior of the ecosystem over time would remain essentially unchanged. We demonstrate situations where the system’s behavior changes dramatically and cannot be captured by a conventional treatment. The complications we explore imply that *short-term descriptions* of what is happening in an evolutionary ecological model can be insufficient and, in fact, misleading, with regard not just to quantitative details but also to qualitative characteristics of ecological dynamics.

Many modeling approaches in mathematical biology which appear distinct at first glance turn out to be describing the same phenomenon with different equations [45, 46, 47]. What matters for our purposes is not so much which technique is chosen, but whether the underlying assumptions do, in fact, apply.

“Mean-field theory” is a term from statistical physics [48, 49] which has been adopted in ecology [50, 51, 52], referring to an approximation in which each component of a system is modeled as experiencing the same environment as any other. This implies that the probability distribution over all possible states of the system factors into a product of probability distributions for individual components. An example in population genetics is the assumption that a population is panmictic. That is, if a new individual in one generation has an equal chance of receiving an allele from any individual in the previous generation, then we can approximate the ecosystem dynamics using only the proportion of that allele, rather than some more complicated representation of the population’s genetic makeup. Modeling evolution of that population as “change in allele frequencies over time” (per, *e.g.*, [45, 53]) is, implicitly, a mean-field approximation [54]. The mean-field approximation is also in force if one postulates that an individual organism interacts with some subset, chosen at random, of the total population, even if the form and effect of interactions within that subset are complicated (as in, *e.g.*, [55, 56]).

It is well known that real species are not necessarily panmictic. However, many treatments which acknowledge this are still mean-field models. The textbook way of incorporating geographical distance into a population-genetic model is to divide the system into N local subpopulations, “islands,” connected via migration [57, 58, 59]. Within each subpopulation, distance is treated as negligible, and organisms are well mixed [44, 60]. This approach makes a simplifying assumption that there is a single distance scale below which panmixia prevails [61], and it relies on well-defined boundaries between panmictic subpopulations which persist over time [60]. Furthermore, the connections among subpopulations are frequently taken to have the topology of a

complete graph, *i.e.*, an organism in one subpopulation can migrate to any other with equal ease [44, 58, 59, 60]. In this case, each of the N subpopulations do experience the same environment, to within one part in N . Thus, the mean-field approximation is in force at the island level, and the island model incorporates spatial extent without incorporating a full treatment of spatial heterogeneity. For real ecosystems [61, 62, 63, 64], one or more of these simplifying assumptions can fail. Long-distance migration is often thought to return a spatial ecosystem to a well-mixed form, but if organisms' migration habits are themselves adaptive, this is not necessarily so [65]. More complicated population structures require more sophisticated mathematical treatments of evolution, a fact which has mathematical consequences, but more importantly has real-world implications for practical issues like the evolution of drug-resistant diseases [66].

Where mean field approximations fail, "higher order" approximations may be employed. Rather than individual organisms or islands, a pair approximation considers pairs of organisms or pairs of spatial regions in average contexts. However, this approximation can also fail when local contexts of groups do not reflect the overall system behavior due to heterogeneity across larger domains. Patches of distinct genetic composition in different parts of a spatial system that are well separated cannot be treated correctly by such approximations. Quantitative analyses confirm this inadequacy. We introduce a new approach to analyzing such approximations by swapping pairs of organisms in a way that preserves the pair description. For spatial systems, such swapping events violate the spatial separation between patches and changes the evolutionary behavior of the system. The swapping method therefore serves as a direct test of the (in)adequacy of the pair approximation. For evolution on random networks of sites that do not embody large spatial distances, the pair approximation can work and the swapping test does not change measures of evolutionary dynamics. However, such networks do not capture important properties of spatial heterogeneity.

As one of the key properties of spatial extent is the propagation of organisms from one part of the space to the other over long distances, we show that important insights can be gained by considering models of percolation. Percolation describes the physical propagation of, *e.g.*, fluids through a random medium. In certain limits the evolutionary behavior of spatial systems can be mapped onto percolation behavior, demonstrating that investigations of such systems which go beyond mean-field or scaling studies are relevant to evolutionary dynamics. This and other advances that go beyond the mean field are necessary to fully describe spatial evolutionary dynamics as they are necessary for the description of many physical systems of spatial extent. The complexities of spatially extended evolutionary dynamical systems beyond the prototypical problem of percolation create new demands and opportunities for advancing our insight into the dynamics of heterogeneous systems and their implications for evolution.

3.2 Model and Methods

We make the issue of spatial heterogeneity concrete by focusing on a specific model of ecological and evolutionary interest. We take a model of hosts and consumers

3 Host–Consumer Evolution by Simulation

interacting on a 2D spatial lattice. Each lattice site can be empty (0), occupied by a host (H) or occupied by a consumer (C). We use the term *consumer* as a general label to encompass parasites, pathogens and predators. Where convenient for examples, we will specialize to one or another of these terminologies. Hosts reproduce into adjacent empty sites with some probability g per site, taken as a constant for all hosts. Consumers reproduce into adjacent sites occupied by hosts, with probability τ per host; sometimes τ is fixed for all consumers, but we also consider cases in which it is a mutable parameter passed from parent to offspring. We will refer to τ as the *transmissibility*. Hosts do not die of natural causes, while consumers perish with probability v per unit time (leaving empty sites behind). Because consumers can only reproduce into sites where hosts live, the effective graph topology of reproductively available sites experienced by the consumers is constantly changing due to their very presence. This makes the ecosystem an *adaptive network*, a system in which the dynamics *of* a network and the dynamics *on* that network can occur at comparable timescales and reciprocally affect one another [67,68,69,70,71]. In this model, dynamics can be highly complex, including spatial cascades of host and consumer reproduction. Even when a quasi-steady-state behavior emerges, as we shall see, it is a consequence of fluctuations over extended space and time intervals.

Several different types of biological interactions can be treated by this modeling framework. Hosts could represent regions inhabited by autotrophs alone, while consumers represent regions containing a mixture of autotrophs and the heterotrophs which predate upon them [72]. Alternatively, host agents could represent healthy organisms, while consumers represent organisms infected with a parasite or pathogen. Thus, host–consumer models are closely related to Susceptible–Infected–Recovered (SIR) models, which are epidemiological models used to understand the spread of a disease through a population. SIR models describe scenarios in which each individual in a network is either susceptible (S) to a pathogen, infected (I) with it, or recovered (R) from it; susceptible nodes can catch the disease from infected neighbors, becoming infected themselves, while nodes which have become infected can recover from the disease and are then resistant against further infection. Susceptible, infected and recovered individuals roughly correspond to hosts, consumers, and empty cells, respectively. An important difference between host–consumer models and epidemiological models concerns the issue of *reinfection*. In the host–consumer model, an empty site left behind by a dead consumer can be reoccupied by another consumer, but only if a host reproduces into it first. Other research has considered models where R[ecovered] individuals can also become I[nfected], with a different (typically lower) probability than S[usceptible] ones, thereby incorporating imperfect immunity into the model [73,74]. The degree of immunity is independent of geography and the environment of the R[ecovered] individual, unlike reoccupation in the host–consumer model. Another application is illustrated by the Amazon molly, *Poecilia formosa*, which is a parthenogenetic species: *P. formosa*, all of which are female, reproduce asexually but require the presence of sperm to carry out egg development. (This kind of sperm-dependent parthenogenesis is also known as gynogenesis.) *P. formosa* are thus dependent on males of other species in the same genus—usually *P. mexicana* or *P. latipinna*—for reproduction. Because *P. formosa* do not incur the cost of sex, they

can outcompete the species on which they rely, thereby possibly depleting the resource they require for survival, *i.e.*, male fish [58, 75]. Thus, hosts could be regions containing sexual organisms, with consumers standing for areas containing both sexual and asexual individuals [58].

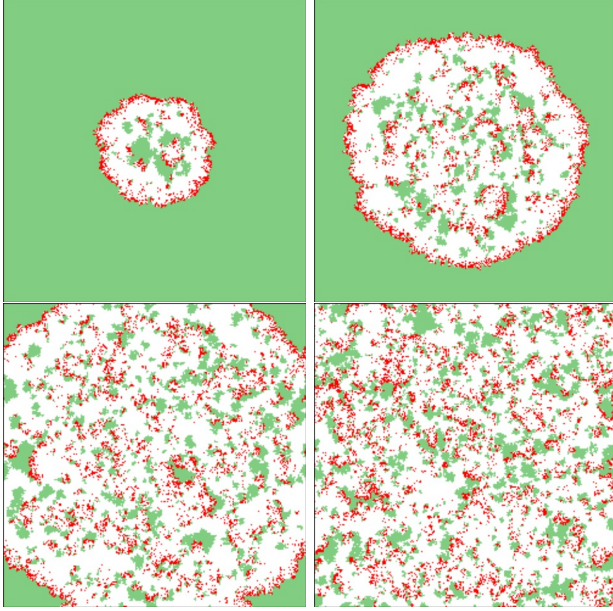


Figure 3.1: Snapshots of a simulated host–consumer ecosystem on a 250×250 lattice, taken at intervals of 100 generations. Consumers are dark gray (red online), hosts are light gray (green online) and empty space is left white. The simulation began with a single consumer at the center of the lattice, which gave rise to an expanding front of consumers. The first image in this sequence shows the state of the ecosystem 100 generations into the simulation. Hosts which survive the consumer wave recolonize the empty sites, leading to pattern formation. Here, the host growth rate is $g = 0.1$, the consumer death rate is $v = 0.2$ and the consumer transmissibility is fixed at $\tau = 0.33$.

This host–consumer model displays waves of colonization, consumption and repopulation. Hosts reproduce into empty sites, and waves of consumers follow, creating new empty regions open for host colonization. Therefore, clusters of hosts arise dynamically [76, 77, 78, 79], a type of pattern formation which can separate regions of the resources available to pathogens into patches without the need for such separation to be inserted manually. Figure 3.1 illustrates a typical example of this effect. This is a specific example of the general phenomenon of pattern formation in nonequilibrium systems [54]. Consumers are *ecosystem engineers* [80, 81, 82, 83, 84, 85] which shape their local environment: an excessively voracious lineage of consumers can deplete the

available resources in its vicinity, causing that lineage to suffer a Malthusian catastrophe [58, 64, 76, 86, 87, 88, 89, 90]. Because the ecology is spatially extended, this catastrophe is a local niche annihilation, rather than a global collapse [91]. A mutant strain with a high transmissibility can successfully invade in the short term but suffer resource depletion in the medium term, meaning that in a population where consumer transmissibilities evolve, averages taken over long numbers of generations yield a moderate value [50, 92]. This implies that an empirical payoff matrix or reproduction ratio will exhibit nontrivial timescale dependence [50, 93, 94].

This model is distinct from another approach to studying evolutionary dynamics in spatial contexts, that of evolutionary game theory. Game-theoretic models of spatially structured populations have been explored at great length. These investigations have found that breakdowns of mean-field approximations are commonplace. However, evolutionary game theory has its own simplifying assumptions. The vast majority of studies consider only two-player games. Population size is usually taken to be constant, and population structure is typically fixed in place. In game-theoretic models, the benefits and costs of different organism behavioral traits are parameters whose values are chosen by the modeler. By contrast, “benefits” and “costs” in host–consumer models are emergent properties which depend on interactions over many generations. Population size is not fixed, and population structure is dynamical: the environment in which different consumer varieties compete changes stochastically, in ways affected by their presence.

3.3 Results

3.3.1 Evolution of Transmissibility

We investigate evolution in the spatial host–consumer ecosystem through simulation and analytic discussion. If the transmissibility τ is made a heritable trait, passed from a consumer to its offspring with some chance of mutation, what effect will natural selection have on the consumer population? Figure 3.2(A) shows the average, minimum and maximum values of the transmissibility τ observed in a population over time. The average τ tends to a quasi-steady-state value dependent on the host growth rate g and the consumer death rate v ; if the simulation is started with τ set to below this value, the average τ will increase, and likewise, the average τ will decrease if the consumer population is initialized with τ over the quasi-steady-state value. Even when the average τ has achieved its quasi-steady-state value, the population displays a wide spread of transmissibilities whose extremes fluctuate over time [72].

In a well-mixed ecosystem, the average τ of the population will tend to 1, maximizing the reproductive rate of the individual consumer. This occurs because each consumer on average experiences the same environment as any other, and thus has the same number of hosts available to reproduce into. A consumer with a higher τ has a higher reproduction rate and therefore evolutionary dominance up to the highest possible value, 1. The observation of a quasi-steady-state value below 1 is an important result. This is the first breakdown of the mean-field approximation, and it indicates

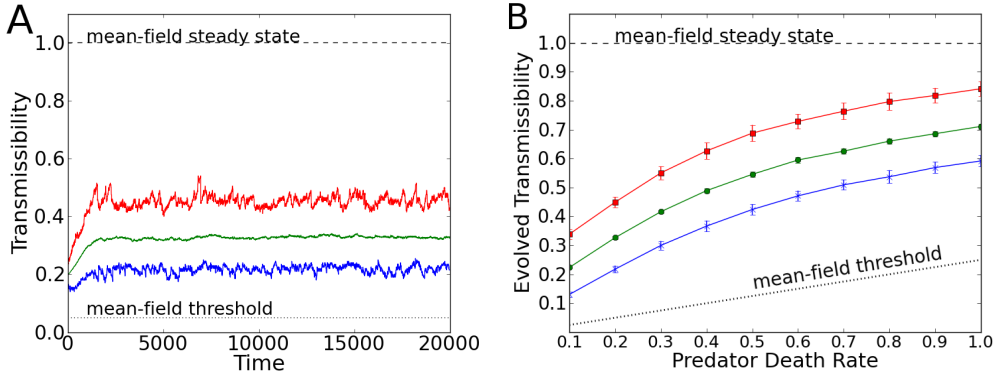


Figure 3.2: **(A)** Minimum (blue), average (green) and maximum (red) transmissibility τ for a consumer population over time, with $g = 0.1$ and $v = 0.2$. (The mutation rate is $\mu = 0.255$ and the step size is $\Delta\tau = 0.005$, as was used in reference [72].) The average τ tends to a quasi-steady-state value dependent on g and v ; if the simulation is started with τ set to below this value, the average τ will increase, and likewise, the average τ will decrease if the consumer population is initialized with τ over the quasi-steady-state value [72]. The horizontal dotted line indicates the threshold value of τ which, in a mean-field model, is the smallest value at which a consumer population can sustain its numbers. The dashed line indicates the value to which τ would trend in a well-mixed ecosystem. **(B)** Minimum, average and maximum τ as a function of v , with $g = 0.1$. The dotted line shows the minimum sustainable τ as predicted by mean-field approximation. Each point is found by averaging over 15,000 timesteps. Error bars indicate one standard deviation.

the inapplicability of traditional assumptions about fitness optimization, with implications for the origins of reproductive restraint, communication-based altruism and social behaviors in general [50, 72, 91, 93, 94, 95].

One can avoid τ tending to 1 in a panmictic system by imposing some extra constraint, such as a tradeoff between transmissibility and lethality, where higher transmissibility becomes impossible due to lethality that prevents transmission. This tradeoff between infectiousness and lethality can be considered as a within-host version of resource overexploitation that here occurs at the population level. Such within-host tradeoffs are difficult to establish empirically in living populations [96, 97]. Often, one lacks pertinent information, such as the functional relationship between pathogen load and disease transmission probability, or the extent to which empirical proxies for pathogen load predict actual host mortality [98]. An empirical observation of low virulence should not by itself be taken as evidence that a tradeoff exists: it may well be that another condition, such as panmixia, fails to obtain. The behavior of spatial

3 Host–Consumer Evolution by Simulation

models makes clear that the relevant scale of the limiting factor is not necessarily within the individual host.

Another difference between spatial and nonspatial host–consumer systems is the rate at which consumers must reproduce in order to sustain their population. One can calculate the minimum sustainable value of τ in the mean-field approximation [99] by balancing the birth and death rates. If the host population is small compared to the total ecosystem size, then the minimum sustainable τ is the value which satisfies $k\tau = v$, where k is the number of neighbors adjacent to a site. For the parameters used in Figure 3.2(A), this value would be 0.05, which is substantially smaller—by a factor of 4—than the lowest τ seen in the evolving spatial population. Consumer populations with τ at the mean-field threshold are not sustainable in the spatial case. This is easily verified by numerical simulations or by using the mean-field equations for the host–consumer dynamics [95, 100, 101]. Stochastic fluctuations suppress the active phase, *i.e.*, the range of parameter values which permit a living consumer population is reduced [99].

To gain insight into this phenomenon, we study the case of fixed τ by means of numerical simulations. We fill the lattice with hosts and inject a single consumer with a τ of our choice; then, we observe how long the descendents of that consumer persist as a function of τ . The consumer population does not persist when τ is either too low or too high. Figure 3.3 shows the probability that a consumer strain will survive for a substantial length of time (2000 generations) after injection into a lattice filled with hosts. This probability is hump-shaped, with an asymmetric plateau bounded above and below by cutoffs.

To understand the upper cutoff visible in Figure 3.3, *i.e.*, the value of τ above which the consumer population again becomes unsustainable, consider the limiting scenario where $g \approx 0$. If hosts do not reproduce into available empty sites, our system reduces to an *epidemic process* which has been studied before [74, 102]. Below the transition point at $\tau = 0.5$, a consumer injected into a lattice of hosts will produce a consumer strain (which we can think of as an infection) which survives for a finite number of generations and then dies out, leaving the lattice filled with hosts (susceptibles) marred by a small patch of empty sites (recovered individuals). Above the transition point, a single consumer gives rise to an expanding wave of consumers which propagates over the lattice, leaving empty sites in its wake, until it consumes all the hosts in the ecosystem. This regime is known as *annular growth*. No finite ecosystem can sustain annular growth indefinitely. If the host growth rate g is made nonzero, then hosts can recolonize sites left empty by the expanding consumer population, opening the possibility of host–consumer coexistence in an ecosystem of dynamically formed and re-formed patches. Figure 3.1 illustrates an example of this phenomenon. This is a specific example of the general phenomenon that, even far from phase transitions, nonequilibrium systems can display *pattern formation* [54].

We can, therefore, interpret the upper cutoff on consumer sustainability as a Malthusian catastrophe due ultimately to the limited amount of available hosts [95]. In physics jargon, this cutoff is a *finite-size effect*. This is the key to understanding what happens when multiple types of consumer are present on the same lattice, and in particular the case we study in the next section, where an invasive consumer variety is introduced to

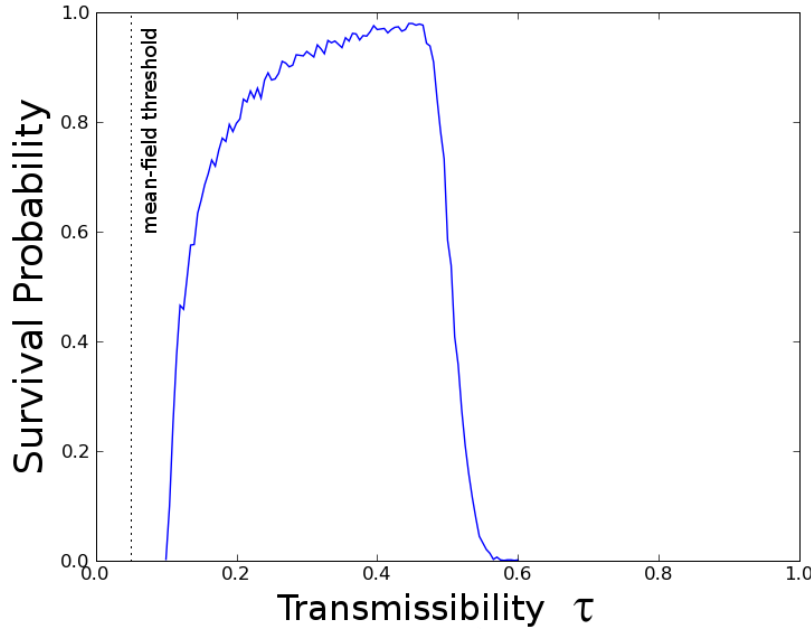


Figure 3.3: Probability of a consumer strain surviving 2000 generations after injection at a single point in a 250×250 lattice filled with hosts. (Computed with 1000 runs per point. Host reproduction rate $g = 0.1$, consumer death rate $v = 0.2$.) Vertical dotted line shows the sustainability threshold found through the mean-field approximation.

an ecosystem where native hosts and consumers have already formed a dynamic patch distribution. The environment experienced by the invasive variety is that formed by the native species, and the “finite size” of the resources available to the invasive variety is not the size of the whole lattice, but that of a local patch [95].

3.3.2 Timescale Dependence of Invasion Success

A key question about an ecological system is whether a new variety of organism, having a different genetic character and phenotypic trait values, can successfully invade a native population. If a mutant consumer strain with fixed transmissibility τ_m can successfully invade a population of transmissibility $\tau_0 < \tau_m$, then we expect the time-averaged value of τ seen in the evolving system to be larger than τ_0 . To investigate this, we simulate scenarios where the native population has τ close to the average value seen in the evolutionary case described above. We then inject a mutant consumer strain with significantly larger τ and study the results. For a typical example, we see from

3 Host–Consumer Evolution by Simulation

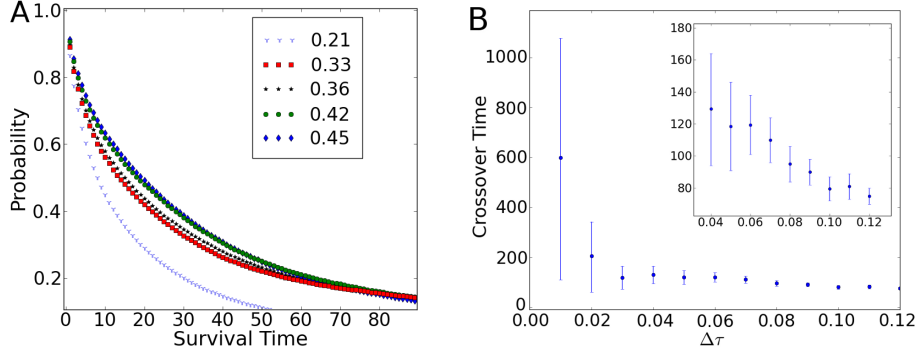


Figure 3.4: **(A)** Survival probability as a function of time for five scenarios: injecting mutants with the same transmissibility as the native consumers, three examples of injecting mutants with transmissibility higher than the τ of the native consumers, and an example of injecting mutants with lower τ than the native population. **(B)** Time intervals during which the survival-probability curves for the native and invasive strains overlap. $\Delta\tau$ indicates the difference between the invasive and native transmissibilities. The closer the mutant trait value is to the resident, the greater the duration of time over which the survival-probability curves for the native and mutant strains overlap. Here, overlap is defined by probabilities being coincident at the 95% confidence level; using other overlap criteria gives qualitatively the same results. Inset: magnified view of the $\Delta\tau \geq 0.04$ region.

Figure 3.2(A) that when $g = 0.1$ and $v = 0.2$, the average τ is approximately 0.33. So, we simulate $\tau_m = 0.45$ mutants entering an ecosystem whose native population has $\tau_0 = 0.33$. Initially, the mutants prosper, but they ultimately fail to invade. As shown in Figure 3.4, the probability of a $\tau_m = 0.45$ strain surviving for tens of generations after injection is larger than that of a $\tau_0 = 0.33$ strain. That is, mutants with the higher τ can out-compete the neutral case. However, after ≈ 74 generations, the survival-probability curves cross. Observed over longer timescales, the mutant strain is less successful than the native variety. This pattern is consistent for $\tau_m > \tau_0$: the average transmissibility seen in the evolutionary case stands up to invasive varieties. This key result manifests the distinctive properties of the spatial structure of the model. The underlying reason for this result is that the mutants encounter the resource limitations imposed by the patchy native population. Over short timescales, the mutant strain enjoys the resources available within the local patch, consuming those resources more rapidly than can be sustained once it encounters the limitations of the local patch size. In this way, the initial generations of the mutant strain “shade” their descendants. Thanks to descendant-shading, short-term prosperity is not a guarantee of medium- or long-term success.

This is to be contrasted with what happens in a well-mixed ecosystem. In the

well-mixed scenario, consumer strains with higher τ successfully invade and displace the native population with a high probability. The invasion success is consistent with the dynamics of a continuously evolving ecosystem. If τ is made an evolvable trait in simulated panmictic systems, the average τ of the population will tend to 1, as predicted by the mean-field analytic proof. There is no difference in a well-mixed scenario between short-term and long-term success. Descendant-shading does not occur in the well-mixed case. This follows from the lack of distinction between local patches and large-scale structure.

One common measure of evolutionary success is the expected relative growth rate of the number of offspring of a mutant individual within a native population, *i.e.*, the relative growth rate of a mutant strain. This rate, known as the *invasion fitness*, is often used to investigate the stability of an evolutionary ecosystem [51, 103, 104]. If the invasion fitness is found to be positive, the native variety is judged to be vulnerable to invasion by the mutant. Conversely, if the invasion fitness is found to be negative, the native variety is deemed to be stable. For the spatial host–consumer ecosystem, this method gives qualitatively incorrect predictions for evolutionary dynamics.

Our investigation builds on earlier work which studied the timescale dependence of fitness indicators in spatial host–consumer ecosystems [93, 94]. In this chapter we have augmented the prior work by considering the survival probability to show the effects of varying τ . We have also more systematically shown the number of generations until dominance of the evolutionary stable strain. In addition, we reported the case of a mutant strain invading a background population, clarifying the conceptual and quantitative results of those earlier works, which considered instead scenarios complicated by multiple ongoing mutations.

3.3.3 Pair Approximations

The inadequacy of mean-field treatments of spatial systems motivates the development of more elaborate mathematical methods. In this section, we review one such methodology, based on augmenting mean-field approximations with successively higher-order correlations, and we test its applicability to our host–consumer spatial model. The numerical variables used in this methodology are probabilities which encode the state of the ecosystem and can change over time. One such variable is, for example, the probability p_a that a lattice site chosen at random contains an organism of type a . Another is p_{ab} , the probability that a randomly-chosen *pair* of neighboring sites will have one member of type a and the other of type b . The change of these quantities over time is usually described by differential equations, for which analysis tools from nonlinear dynamics are available [51, 76, 95, 104, 105, 106, 107].

The importance of the joint probabilities p_{ab} is that they reflect correlations which mean-field approximations neglect. To understand the relevance of the joint probabilities p_{ab} , consider a scenario where an invasive mutant variety forms a spatial cluster near its point of entry. Let p_M be the probability that a lattice site chosen at random contains a mutant-type organism, and let p_{MM} denote the probability that a pair of neighboring sites chosen at random will both be occupied by mutant-type organisms. Then the average density of invasive mutants in the ecosystem, p_M , will be low, while

3 Host–Consumer Evolution by Simulation

the conditional probability that a neighbor of an invasive individual will also be of the invasive type, $q_{M|M} = p_{MM}/p_M$, will be significantly higher. (It is typical in theoretical spatial ecology to denote conditional probabilities with q , rather than p [108].) A discrepancy between the conditional probability $q_{a|b}$ and the overall probability p_a can persist when the ecosystem has settled into a quasi-steady-state behavior, and is then an indicator of spatial pattern formation.

Applying this idea to the spatial host–consumer model, let p_C be the probability that a lattice site chosen at random contains a consumer, and let $q_{C|C}$ denote the conditional probability that lattice site adjacent to a consumer will also be occupied by a consumer. Figure 3.5(A) shows p_C and $q_{C|C}$ measured during the course of numerical simulations. In a well-mixed scenario (where we expect the mean-field approximation to be applicable), the average consumer density p_C and the consumer–consumer pairwise correlation $q_{C|C}$ are essentially equal over time. In the spatial lattice scenario, p_C and $q_{C|C}$ are noticeably different.

Treating the correlations $q_{a|b}$ as not wholly determined by the probabilities p_a is a way of allowing spatial heterogeneity to enter an analytical model. Whether it is a *sufficient* extension in any particular circumstance is not, *a priori*, obvious. Typically, the differential equations for the pair probabilities p_{ab} depend on triplet probabilities p_{abc} , which depend upon quadruplet probabilities and so forth. The standard procedure is to truncate this hierarchy at some level, a technique known as *moment closure* [70, 103, 105, 109, 110]. Moment closures constitute a series of approximations of increasing intricacy [111, 112]. The simplest moment closure is the mean field approximation; going beyond the mean field to include second-order correlations but neglecting correlations of third and higher order constitutes a *pair approximation*. These approximations do not incorporate all of the information about spatial structure which may be necessary to account for real-world ecological effects [103].

3.3.4 Organism Swapping

Several factors have been identified which undermine pair approximations [70, 72, 95, 100, 101, 103, 106, 113, 114, 115]. In our model, we can directly test the efficacy of pair approximations in a completely general way. The key idea is to transplant individuals in such a way that the variables used in the moment-closure analytical treatment remain unchanged. At each timestep, we look through the ecosystem for isolated consumers, that is, for individual consumers surrounded only by a specified number of hosts and empty sites. We can exchange these individuals without affecting the pairwise correlations. For example, if we find a native-type consumer adjacent to three hosts and one empty site, we can swap it with an invasive-type consumer also adjacent to three hosts and one empty site. We can also exchange isolated pairs of consumers in the same way. The variables used in the moment-closure treatment remain the same. Were the moment-closure treatment valid, we would expect the dynamics to remain unchanged when we perform such exchanges.

When we perform the simulation, however, swapping strongly affects the dynamics. With this type of swapping in effect, mutants with higher τ can invade a native population with lower τ . In one typical simulation run with a native τ of 0.33 and an

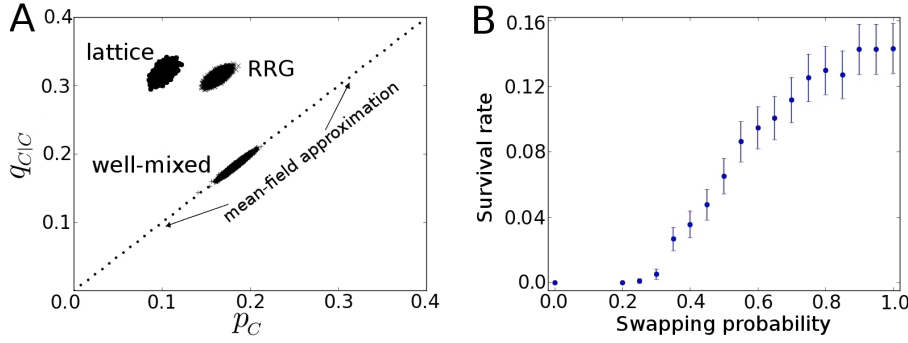


Figure 3.5: **(A)** Pairwise conditional probability $q_{C|C}$ plotted against the average density of consumers, p_C , for three variations on the host–consumer model: a well-mixed case in which mean-field theory is applicable, a random regular graph (in which each site has exactly four neighbors) and a 2D square lattice. The dotted line, $p_C = q_{C|C}$, indicates the mean-field approximation. 10^4 timesteps were computed for each case. The well-mixed case is simulated by dynamically rewiring sites at each time step, precluding the generation of spatial heterogeneity; consequently, the pairwise correlation $q_{C|C}$ is within statistical variation equal to p_C ($R^2 = 0.953$). A random regular graph (RRG) with random but static connections does develop spatial heterogeneity so that $q_{C|C}$ is not the same as p_C ($R^2 = 0.581$). The discrepancy is even stronger in the lattice case ($R^2 = 0.304$). **(B)** Success rate of invasive mutant strains as a function of swapping probability. Voracious mutant strains with $\tau = 0.45$ are introduced into a lattice ecosystem defined by a host growth rate of $g = 0.1$, a consumer death rate $v = 0.2$ (the same for both consumer varieties), and a native consumer transmissibility of $\tau = 0.33$. Average success rates are found by simulating 2000 invasions per value of the swapping probability parameter; error bars indicate 95%-confidence intervals. Increasing the fraction of possible swaps which are actually performed makes the voracious invasive strain more likely to take over the ecosystem.

invasive τ of 0.45, the invasive strain succeeded in 1,425 of 10,000 injections. Without swapping, the number of successful invasions is *zero*.

Swapping can be considered as creating a new ecosystem model with the same moment-closure treatment as that of the original. The behavior of invasive strains is different, because transplanting organisms allows invasive varieties to evade localized Malthusian catastrophes. Swapping opens the ecosystem up to invasive strains, since, in essence, it removes individuals from the “scene of the crimes” committed by their ancestors.

This type of swapping is, to our knowledge, a new test of moment-closure validity. Randomized exchanges have been incorporated into computational ecology simulations

3 Host–Consumer Evolution by Simulation

for different purposes. For example, research on dispersal rates in an island model shuffled individuals in such a way that the population size of each island was held constant [116].

If, instead of performing every permissible swap, we transplant organisms with some probability between 0 and 1, we can interpolate between the limit of no swapping, where invasions always fail, and the case where pair approximation is most applicable and invasions succeed significantly often. The results are shown in Figure 3.5(B) and indicate that the impact of swapping becomes detectable at a probability of ≈ 0.25 and effectively saturates at a probability of ≈ 0.9 .

Our swapping method allows us to test the significance of complications which can undermine pair approximation techniques or make them impractical to apply, several of which have been identified. First, introducing *mutation* into a game-theoretic dynamical system can make pair approximation treatments of that system give inaccurate predictions [106, 114].

Second, when the evolving population has a network structure, the presence of *short loops* in the network often makes pair approximations fail [113]. For example, in a triangular lattice, one can take a walk of three steps and return to one’s starting point, whereas on a hexagonal lattice, the shortest closed circuit is six steps long. A pair approximation can work well for a dynamical system defined on the hexagonal lattice but fail when the same dynamics are played out on a triangular one. This happens because the short loops provide opportunities for contact which the coarse-graining necessary for a pair approximation will miss. (We will study this point in more detail in Chapter 4.) This effect is amplified in adaptive network models, where the underlying network changes dynamically in response to the population living upon it. In such cases, even extending the moment closure to the triplet level brings little improvement [70].

Third, *fluctuating population sizes* make pair approximations significantly more cumbersome to construct, leading to systems of differential equations which are too intricate to be significantly illuminating. In a game-theoretic model where a lattice is completely filled at all times with cooperators and defectors, there is one independent population density variable and three types of pairs. By contrast, in an ecological model where two consumer varieties are competing within an adaptive network of hosts, a pair approximation requires nine independent variables [95, 100, 101]. Modeling phenomena of biological interest can easily increase the complexity still more. For example, if organism behavior changes in response to social signals [72], the number of possible states per site, and thus the number of dynamical variables in a pair approximation treatment, increases further.

Fourth, the pair-approximation philosophy of averaging over all pairs in the system impedes the incorporation of *environmental heterogeneities*. These include biologically crucial factors like variable organism mobility, background toxicity or other localized “costs of living,” and resource availability [115].

Finally, *dynamical pattern formation* creates spatial arrangements which the pair approximation does not describe [103]. This can be thought of as the ecosystem generating its own environmental heterogeneities.

3.3.5 Effect of Substrate Topology

It is instructive to compare the spatial lattice ecosystem with the host–consumer model defined on a random regular graph (RRG). In an RRG, each node has the same number of neighbors, as they do in a lattice network, but the connections are otherwise random. RRGs have been used as approximations to incorporate the effects of spatial extent into population models, as they make for more tractable mathematical treatments, although they are typically less realistic than spatial lattices [117]. The network structure is set at the beginning of a simulation and does not change over time. The important aspect of this network as compared to the spatial case is that there exist short paths of links that couple all nodes of the network. This is quite different from the spatial case, where strains in one part of the network cannot reach another in only a few generations due to the need to traverse large numbers of spatially local links.

When we simulate our host–consumer ecosystem on an RRG, we find that an invasive consumer strain with higher transmissibility τ can out-compete and overwhelm a native consumer population with lower τ . In one typical simulation run, using the native and invasive τ values of 0.33 and 0.45 respectively, 2,233 out of 10,000 invasions were successful, whereas on the lattice *no* invasion succeeded using the same parameters. Thus, the RRG does not capture the essential features of the spatial scenario. In particular, our results show that the RRG case is more like the well-mixed case than the spatial lattice, as far as stability against invasion is concerned.

Our swapping test provides insight into the utility of the pair approximation, which can be effective for the RRG even though it is not for the spatial case. Consider the pairwise correlation value $q_{C|C}$, which would be a variable for a pair approximation treatment. On an RRG, the underlying network topology provides enough locality that p_C and $q_{C|C}$ are unequal, distinct from the well mixed case as shown in Figure 3.5(A). This means that the pair approximation is nontrivial for the RRG as it incorporates the difference between $q_{C|C}$ and p_C , which would not be contained in a mean-field treatment. We can also implement swapping on the RRG, where invasions can succeed without it; as expected, swapping does not affect the success rate on the RRG. With 10,000 simulated invasions for each case, the 95%-confidence interval for the difference in success rates between full swapping and none is 0.004 ± 0.01 . Thus, the pair approximation may be successful in this network topology. However, this does not mean that the RRG or the pair approximation capture the full significance of a spatial system, because the RRG network does not embody essential properties of spatial extent—separation by potentially large distances.

3.3.6 Percolation

In order to obtain quantitatively or even qualitatively correct predictions for spatial host–consumer evolutionary dynamics, different approaches are needed. Having encountered the limitations of moment closures, we now demonstrate a change of perspective which yields quantitatively useful results. In certain situations, the process of pathogen propagation through the host population distributed in space can be mapped onto a *percolation* problem. A topic widely investigated in mathematics, percolation

3 Host–Consumer Evolution by Simulation

theory deals with movement through a matrix of randomly placed obstacles. A prototypical percolation problem is a fluid flowing downhill through a regular lattice of channels, with some of the lattice junction points blocked at random. The key parameter is the fraction of blocked junction points. If this fraction is larger than a certain threshold value, the fluid will be contained in a limited part of the system. However, if the blocking fraction is below the threshold, the fluid can percolate arbitrarily far from its starting point. This is a *phase transition*, a shift from one regime of behavior to another, in this case between a phase in which fluid flow can continue indefinitely and one in which flow always halts. Similar issues arise when a pathogen propagates by cross-infection through a set of spatially arranged hosts. Sufficiently many hosts in mutual contact are required for the pathogen to propagate successfully. Pathogen strains therefore survive or die out over time depending on whether percolation is or is not possible [52, 99, 102, 118, 119, 120].

One important goal of studying host–pathogen models is knowing the pathogen properties that enable its survival in a population, or equivalently what prevents it from persisting in a population. The growth rate of a pathogen in a population can be an important public health concern. We therefore focus on analyzing the minimum value of the transmissibility that enables a pathogen population to persist, and the growth dynamics of population sizes near that transition.

Of essential importance to the quantitative theoretical and empirical analysis is the recognition that infected population growth can be described by power laws $n \sim t^z$, with an exponent that differs from that of the mean field. Identifying the value of the power z is important to practical projections of the number of infected individuals. The initial growth curve of infected populations can be correctly extrapolated if the exponent is known, guiding public health responses. Knowing what impediments are needed to prevent further propagation can even better guide public health intervention strategies.

We show in Figure 3.6 the results of numerical simulations which indicates that the consumer extinction transition, when the transmissibility τ becomes just large enough that the consumer population sustains itself, lies in the *directed percolation universality class* [74, 121, 122, 123, 124]. A similar result has been reported for related models [99, 102], consistent with those models being in the same universality class. The directed percolation universality class is a large set of models, all of which exhibit a phase transition between two regimes of behavior, and all of which behave in essentially the same way near their respective transition points. The scenario of fluid flow through a random medium considered above is a classic example of a directed percolation-class model, but many others exist as well [74, 121]. The *critical exponents* describe how properties of the modeled system vary over time or as a function of how far the control parameter is from the critical point. They are the same for all systems in the universality class. Other universality classes exist as well, with different classes having different quantitative values for the critical exponents. Identifying the universality class a system belongs to enables us to study a complicated phenomenon by examining a simpler representative of its class instead. This is convenient, because regions of parameter space near phase transitions are precisely where mean-field and moment-closure approximations are least reliable, even for short-timescale modeling. Near the

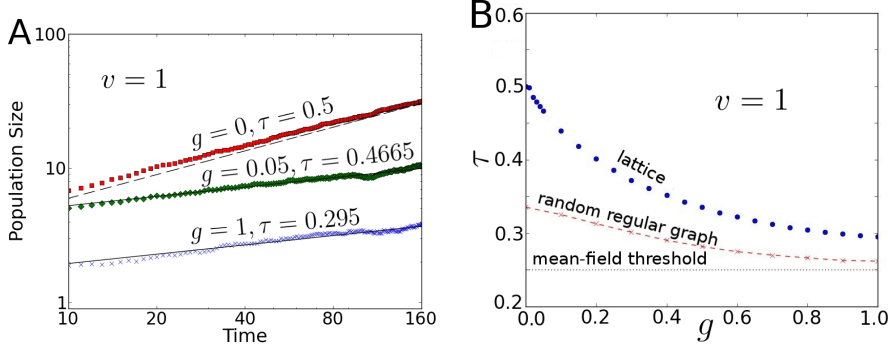


Figure 3.6: **(A)** Population size as a function of time, averaged over 10^3 simulation runs, for τ values near the transition points at $g = 0$ and $g = 0.05$ on the spatial lattice, with $v = 1$. Dashed and solid lines indicate the population growth for systems at dynamic percolation and directed percolation transitions respectively, showing that these transitions have the characteristic properties of those universality classes. **(B)** Critical τ for the host–consumer ecosystem with $v = 1$. The transition line crosses over from the dynamic percolation universality class at $g = 0$ to directed percolation between $g = 0.015$ and $g = 0.02$. Red Xs indicate the transition curve for the host–consumer dynamics on a random regular graph (RRG) of uniform degree 4; the dashed line connecting them is to guide the eye. The RRG transition is neither directed percolation nor dynamic percolation.

phase transition, stochastic fluctuations create dynamical patterns with a wide range of sizes. In Figure 3.6(A), we see that percolation theory gives quantitatively correct predictions for the growth of consumer population sizes in the spatial host–consumer model.

We can understand the $g = 0$ and $g = 1$ extremes by mapping the host–consumer model onto other stochastic models for which exact or approximate results are available. When $g = 0$, the host–consumer model maps onto the SIR epidemic process [74]. In turn the SIR model on the square lattice can be understood in terms of bond percolation on the square lattice [125], for which the transition point is known exactly [126, 127]. We can therefore predict analytically that the critical τ on the square lattice is 0.5. Percolation theory also gives a prediction for the critical τ on an RRG: it should be approximately $1/3$ [128]. These both match the simulation results seen in Figure 3.6(B).

In contrast, when $g = 1$, empty sites are filled as quickly as possible, so the behavior of the host–consumer model should resemble that of an epidemic model with only Susceptible and Infected sites. In this case, the transition point of the epidemic model on the square lattice is only known numerically [125]. The numerical value, ≈ 0.29 , does agree with the critical τ found by simulating the host–consumer model at $g = 1$.

3 Host–Consumer Evolution by Simulation

Thus, in the limiting cases of $g = 0$ and $g = 1$, the host–consumer model is roughly equivalent to the SIR and SIS epidemic models. However, a host–consumer model with $0 < g < 1$ has dynamical behavior distinct from an epidemic model which allows reinfection of Recovered sites. The key difference is that reoccupying an empty site with a consumer requires prior recolonization by a host, whereas the vulnerability of a R[ecovered] individual to becoming I[nfected] is defined as an intrinsic property of the R[ecovered] type. This changes the role of ecology: both models incorporate space, but the effect of spatial extent is different. This manifests as a change in the shape of the critical-threshold curve, as well as a change in universality class [74, 125].

Furthermore, when we use an RRG topology instead, comparing host–consumer dynamics at $g = 1$ with an SIS epidemic model reveals their transitions to take place at different thresholds. For the host–consumer model, the critical τ on an RRG is approximately 0.2615, while the SIS threshold is approximately $1/3$ [129, 130, 131].

The physics analysis of percolation behavior near the transition point maps directly onto the critical public health problem of the growth of infected populations, and more generally onto the dynamics of evolutionary systems. For these systems the mean field treatment fails and the standard transmission of infectious diseases in a population need not apply. Applications to real world systems must accommodate the actual network of connectivity. This network can also be modified by intervention strategies.

3.3.7 Patch Size and Structure

Since it is the size of a host patch which determines the amount of resources available for consumers, we now investigate patch sizes in detail. One way to test if the host patches have a characteristic size is to take snapshots of the dynamics in its quasi-steady-state regime (*e.g.*, the fourth panel of Figure 3.1) and compute its *auto-correlation*. We can do this by running the snapshot through a filter that produces a binary matrix whose entries are 1 in locations occupied by hosts and 0 otherwise. Applying FFTs and the Wiener–Khinchin theorem then yields a 2D autocorrelation matrix, which by averaging we can collapse to a function of distance. If this autocorrelation function has a characteristic distance scale—for example, if it decays exponentially—then we can use that distance as the *correlation length*. We present examples confirming this in Figure 3.7.

Next, Figure 3.8 summarizes the results of this procedure for host–host correlations as we vary τ , holding the other parameters fixed at $g = 0.1$, $v = 0.2$. We see that the correlation length of the host distribution, which we can regard as the characteristic size of host patches, increases with τ . By choosing a different filter, we can apply the same procedure to find the characteristic size of empty regions. The right panel of Figure 3.8 summarizes the results.

We consider the following scenario: a native population of consumers, with transmissibility τ_{nat} , is dynamically forming and re-forming host patches. Into this ecosystem, a mutant consumer with transmissibility τ_{mut} is introduced. Call the characteristic size of host patches ξ_H and the characteristic separation between them ξ_0 . Both of these length scales will depend on g , v and τ . The length $\xi_H(g, v, \tau = \tau_{\text{mut}})$ is the typical size of host patches in which the mutant variety “expects” to live. If this length

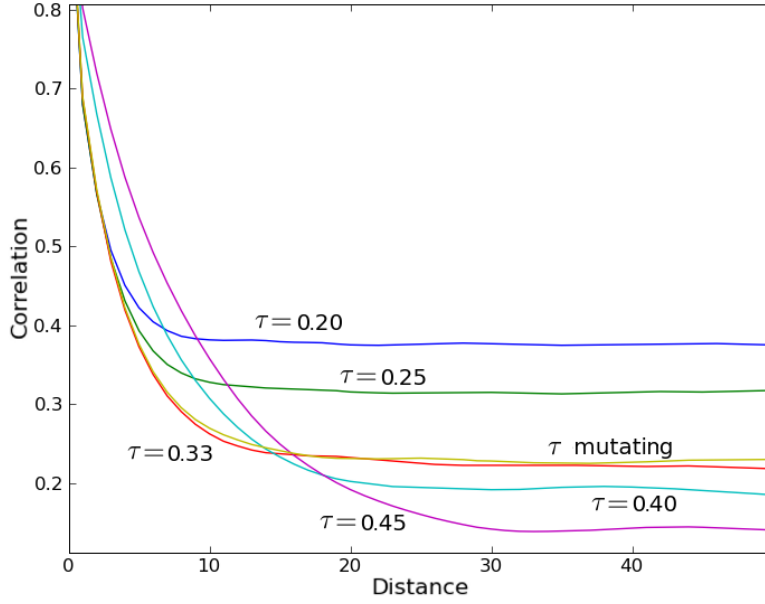


Figure 3.7: The decay of host-host autocorrelation as a function of distance, for different choices of τ (fixing $v = 0.2$, $g = 0.1$). Note that the correlation curve for $\tau = 0.33$ essentially coincides with the curve found when τ is a mutable trait.

is too much larger than $\xi_H(g, v, \tau = \tau_{\text{nat}})$, the size of the patches which exist due to the native population, then the mutant variety is likely to suffer a Malthusian catastrophe.

When we find the correlation lengths via simulations, as seen in Figures 3.8 and 3.9, we see that both ξ_H and ξ_0 increase with τ . Assuming that the deleterious effect of “expecting” larger host patches grows with the discrepancy between “expected” and actual patch sizes, the distribution of τ in the consumer population will accumulate in the region where the ξ_H curve starts to take off.

The missing piece is an *analytical* expression for ξ_H and ξ_0 in terms of the parameters g , v and τ , when all these parameters are fixed. What is the functional form of

$$\xi_H = \xi_H(g, v, \tau) \text{ and } \xi_0 = \xi_0(g, v, \tau) ? \quad (3.1)$$

Figure 3.8 suggests that both correlation lengths depend roughly exponentially on τ , but the offsets and growth rates will depend on g and v .

One way in which we might try to estimate the length scale ξ_H and see, at least qualitatively, how it depends on the system parameters, is to use a moment closure. This method turns out not to work, and we now investigate why. Consider a square

3 Host–Consumer Evolution by Simulation

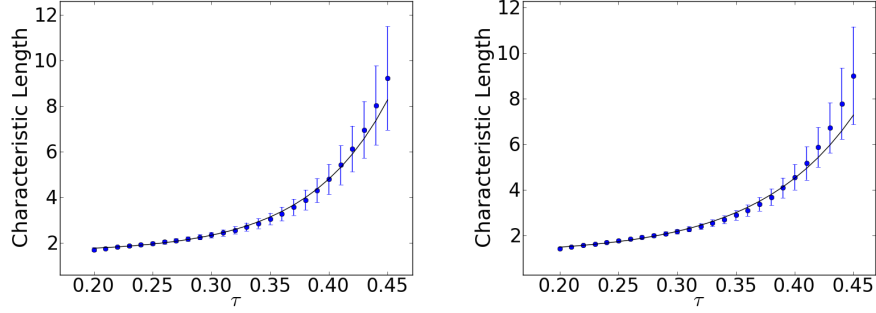


Figure 3.8: **(Left)** Correlation length of hosts for different values of τ ($g = 0.1$, $v = 0.2$). The solid line is the curve $\xi_H = \alpha \times \exp(\tau/\beta) + \gamma$, where the parameters found by least-squares fitting are $\alpha = 0.009 \pm 0.002$, $\beta = 0.068 \pm 0.002$ and $\gamma = 1.59 \pm 0.02$. **(Right)** Correlation length of empty space under the same conditions. The solid line is the curve $\xi_0 = \alpha \times \exp(\tau/\beta) + \gamma$, with $\alpha = 0.026 \pm 0.007$, $\beta = 0.082 \pm 0.004$ and $\gamma = 1.20 \pm 0.05$.

lattice of size $L \times L$, and assume for simplicity that it contains some number K of host patches, all of a comparable size which we take to be the characteristic length ξ_H . The population density of hosts is

$$p_H = \frac{K\xi_H^2}{L^2}. \quad (3.2)$$

The quantity $q_{H|H}$ used in a pair approximation denotes the probability that a randomly chosen neighbor site of a randomly chosen host will also contain a host. If we imagine a patch of hosts in an otherwise empty lattice (for convenience, suppose the patch is roughly circular), then sites in the interior of the patch have all their neighboring lattice sites occupied by hosts, while sites on the edge have some empty space adjacent. Only the sites on the perimeter can be responsible for decreasing $q_{H|H}$ below unity. Writing P for perimeter and A for area, we have

$$q_{H|H} = 1 - \frac{KfP}{A}, \quad (3.3)$$

where f is a constant we can take to be roughly one-half. If, for simplicity, we assume ordinary Euclidean scaling of perimeters and areas,

$$P \sim \xi_H, \quad A \sim \xi_H^2, \quad (3.4)$$

then we have that

$$q_{H|H} = 1 - \frac{Kf'}{\xi_H}. \quad (3.5)$$

We have two quantities, p_H and $q_{H|H}$, which we can compute by numerically integrat-

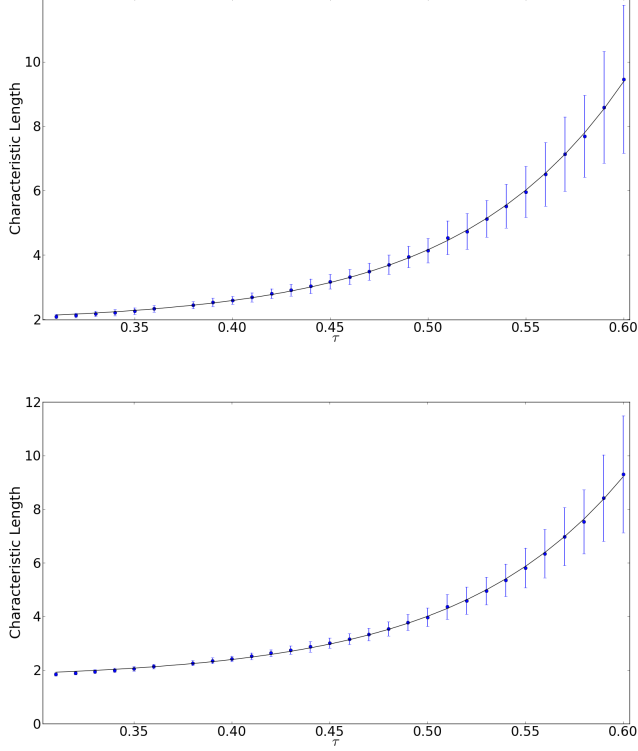


Figure 3.9: **(Top)** Host-host correlation length ξ_H for different values of τ ($g = 0.1$, $v = 0.4$). The solid line is the curve $\xi_H = \alpha \times \exp(\tau/\beta) + \gamma$, where the parameters found by least-squares fitting are $\alpha = 0.0089 \pm 0.0005$, $\beta = 0.0893 \pm 0.0009$ and $\gamma = 1.81 \pm 0.01$. **(Bottom)** Void-void correlation length ξ_0 under the same conditions. Here, $\alpha = 0.014 \pm 0.001$, $\beta = 0.096 \pm 0.002$ and $\gamma = 1.51 \pm 0.02$.

ing a set of coupled differential equations. And we have two unknowns: the patch size ξ_H and the number of patches K . We solve for ξ_H , obtaining

$$\xi_H^3 = f' L^2 \frac{p_H}{1 - q_{H|H}}. \quad (3.6)$$

As the conditional probability $q_{H|H}$ approaches unity, the patch size increases.

Unfortunately, when we compute p_H and $q_{H|H}$ by numerically iterating the dynamical equations (see [95, 100, 101]), we find that the inferred ξ_H value *decreases* with τ . The direction of the change, as predicted by pair approximation, is incorrect! This is a sign that the pair approximation is losing important information about larger-scale

3 Host–Consumer Evolution by Simulation

structures.

Another possible way to predict how ξ_H and ξ_0 depend on g , v and τ comes from the theory of spatial stochastic processes. For some models, we can derive an analytical expression for the power spectrum of fluctuations, as a function of both spatial and temporal oscillation frequencies. A peak in this power spectrum indicates a characteristic length or time scale [132, 133, 134]. Because the host–consumer lattice model differs in a few significant ways from those treated by these methods before, we defer a detailed exploration of this idea until a later chapter. The upshot appears to be that the characteristic length scale increases in the proper direction, that is, with increasing τ , although thanks to the differences hinted at, the concavity may not be correct.

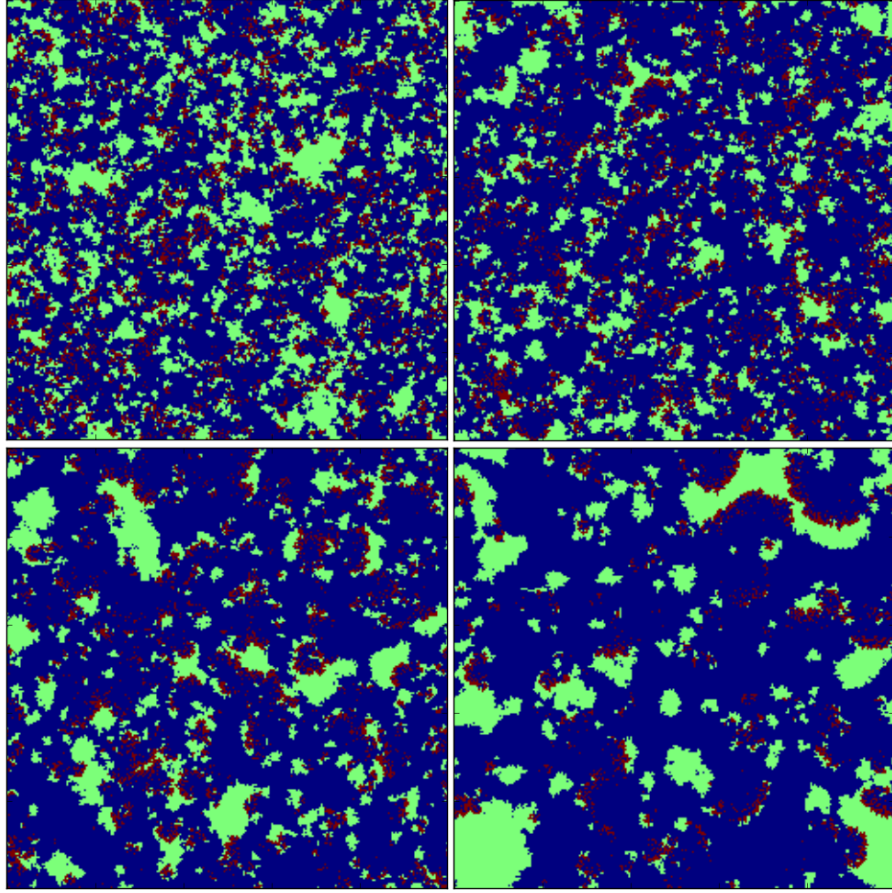


Figure 3.10: Typical frames from host–consumer simulations run with different values of the transmissibility τ (the other parameters are fixed at $g = 0.1$, $v = 0.2$). **(Top left)** $\tau = 0.30$; **(top right)** $\tau = 0.35$; **(bottom left)** $\tau = 0.40$; **(bottom right)** $\tau = 0.45$.

At this point, we might be concerned that computing a correlation length in this fashion could be smearing over too much. If there are many small patches and a few large ones (and Figure 3.10 makes this look plausible), then does reducing the pattern to a single correlation length lose any important information? To cross-check the idea that host patch size controls the evolved transmissibility, we therefore measure patch size in another way.

The most straightforward way to assign a size to a patch is to count the number of lattice sites it contains. This has the advantage that while we are computing it, we can also count the number of sites in the patch which border on empty space. So, we can see how the perimeters of host patches relate to their areas. If host patches were Euclidean circles, then the perimeter would scale as the area to the one-half power. On the other hand, a patch which consists of a single site in a discrete lattice is all boundary: its perimeter and its area are equal. Our patches live on a square lattice, so a cluster of hosts must contain at least five individuals for its area to be greater than its perimeter. The more hosts are contained within the cluster, the more likely the cluster is to have an interior. We therefore expect a crossover: below some value of the area, the maximum observed perimeter will be equal to the area. Above the crossover point, the maximum observed perimeter will grow more slowly. Noting that the patch dynamics are stochastic and their edges irregular, we hypothesize that this growth will be faster than the area-to-the-0.5 obtained for Euclidean circles.

Figure 3.11 bears this out. When we plot the perimeters of host clusters against their areas, we see a crossover at an area of ≈ 30 lattice sites (for $g = 0.1$, $v = 0.2$). Furthermore, the increase of perimeter with area above the crossover point is faster than the square root of the area.

Pascual *et al.* [135] study the cluster-size distribution for prey in a predator-prey model which is similar to our host-consumer ecosystem. However, they do not find a crossover: instead, cluster perimeter scales smoothly and just barely sublinearly with cluster area, across the whole range of observed areas. This may be due to an extra mixing effect which their model includes and ours does not, an effect which tends to bring more of a cluster to its perimeter.

Having measured the host patch sizes, we can investigate the size distribution. Figure 3.12 shows the numbers of patches observed at different sizes. The fall-off of frequency with area is faster than an inverse-area relationship, though not uniform. At first blush, an appropriate characterization would be a power-law decay with an exponential cutoff. Because this distribution is fairly broad, we use a *percentile* to characterize it: we find the value of the area such that 99% of the host patches are that size or smaller. This is indicated by the vertical dashed line in Figure 3.12.

We see the results of repeating this calculation across a range of τ values in Figure 3.13. The *minimum* of the 99th-percentile curve gives the value of the transmissibility which evolves when τ is a mutable trait!

We can understand this, heuristically, using much the same argument we made above. A consumer strain which “expects” that large host patches are available will fare poorly if they are not. When τ is evolvable, the τ distribution of the consumer population will tend to concentrate around the minimum of the 99th-percentile curve. In the short term, local subpopulations with higher τ can blossom, causing occasional

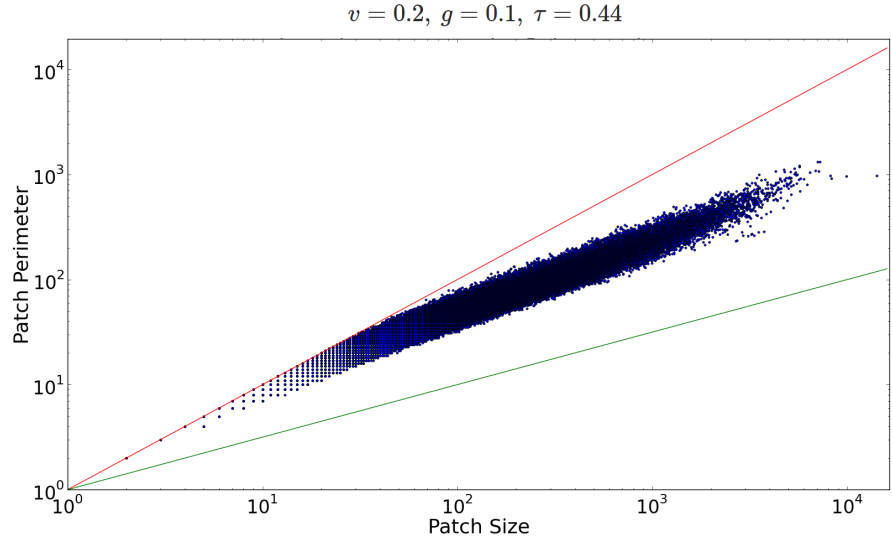


Figure 3.11: Perimeters of host patches, plotted against their areas ($g = 0.1$, $v = 0.2$, $\tau = 0.44$). The upper (red) straight line indicates linear growth, and the lower (green) line indicates square-root growth. Note the crossover at area ≈ 30 lattice sites. Below the crossover point, the maximum observed perimeter toes the line of linear growth, indicating that host patches exist which have no interior.

upward shifts in the average τ . We expect, therefore, that the average τ in the evolving population will lie somewhat to the right of the curve’s minimum point.

The hypothesis that these curves can be well approximated by exponentially truncated power laws can be validated by standard curve-fitting techniques [136]. Applying methods suitable to power-law analysis reveal that for all τ values, the number of patches decays with roughly the 1.5 power of the patch area. The location of the cutoff determined by curve-fitting follows, unsurprisingly, the relationship seen with the 99th-percentile areas (Figure 3.13). One should take care, however, when applying these statistical methods and interpreting the results, as they presume an independence among data points which may or may not be applicable here. If it is the case, for example, that large patches are typically accompanied by smaller patches nearby, then the assumption of statistical independence would be invalid.

The qualitative shape of the area/abundance curve, that is, the power-law dependence with an exponential cutoff, is also seen in the results of *coagulation and fragmentation processes* [137, 138]. Furthermore, such distributions are known to be *maximum entropy* distributions: they arise when one maximizes the Shannon entropy (which we will discuss in Chapter 5), subject to a certain type of constraint [139]. This suggests that the general functional form is rather robust.

It is a good idea to try and understand the shape of the 99th-percentile curves (Figure 3.13) heuristically, since the minima of those curves indicate the values to which τ will evolve. Suppose, for concreteness, that the abundance as a function of area is a power-law decay with an exponential cutoff. Why should the cutoff decrease and then increase? Can we think of this in terms of countervailing influences?

First, we consider the low- τ regime. When τ is small, the cutoff should be large, because the system is near a critical point, and critical points mean power laws. The closer we move towards criticality, the better the quality of the power law, and the less noticeable any cutoff will be. Therefore, as we increase τ , we move away from criticality, and the deviation from a clean power law becomes more severe. This qualitatively explains the decreasing part of the curve in Figure 3.13.

What about when τ is large—say, when it is larger than the value to which it would evolve when mutation is present? Here, the theory of critical points is no longer pertinent, and we find guidance instead in the study of coagulation and fragmentation processes. Numerical simulations indicate that for each value of v and g , there is a region along the τ axis where the *total host population size* does not strongly depend upon τ (see Figure 3.14). That is, when τ is large, we can adjust it and the host population size P will not change too much in response. This suggests that we can construct a simplified model in which P is a parameter that we can vary independently of the consumer transmissibility.

Gueron and Levin develop a model of group-size dynamics wherein a population of fixed total size is divided into groups that can merge and split stochastically [137]. The rates of merging and splitting are taken to be functions of the group sizes. If x and y are the sizes of two groups, then the probability per unit time that those groups will merge is $m(x, y)$. Similarly, the probability per unit time that a group of size x will fission into fragments of sizes y and $x - y$ is $s(x, y)$. A convenient choice of functional form that allows some analytical solutions is to take

$$m(x, y) = \mu a(x) a(y), \quad (3.7)$$

where $a(x)$ is an increasing function of x and μ is a rate parameter. (For example, the probability of merging might increase with the surface area of each group, implying a power-law dependence on the group population.) The splitting rate is taken to be

$$s(x, y) = 2\sigma a(x). \quad (3.8)$$

This stochastic process has a stationary solution: the distribution of patch sizes follows the truncated decay

$$f(x) = \frac{2\sigma}{\mu} \left(\frac{1}{a(x)} \right) e^{-x/x_c}. \quad (3.9)$$

The cutoff x_c is fixed by the total population size:

$$P = \frac{2\sigma}{\mu} \int_0^\infty dz \frac{z}{a(z)} e^{-z/x_c}. \quad (3.10)$$

3 Host–Consumer Evolution by Simulation

In our host–consumer ecosystem, host patches split apart because they are eaten into by consumers. They can merge together on their own, but their fission requires consumption. Very crudely speaking, the splitting rate should increase with the consumer population density. (In a slightly more refined approximation, we could say that the splitting rate should increase with the contact probability $q_{H|C}$.) This corresponds, in the Gueron–Levin model, to increasing σ . If we imagine that P and μ are fixed, increasing σ must be balanced by decreasing the value of the integral. Likewise, lower consumer density corresponds to lower σ and thus a larger value for the integral. The only way to change the value of the integral, if the function $a(z)$ is a given, is to alter the cutoff x_c . To obtain a larger value, we move the cutoff farther out.

We have, therefore, that if the host density is *constant* with respect to τ but the consumer density *falls*, then the cutoff should be *larger*. When we measure the densities by numerical simulation, these are the trends we find at larger τ , and so the upswing fits neatly into our picture.

Figures 3.14 and 3.15 together indicate that while pair approximation itself fails to capture the eco-evolutionary dynamics of the spatial system, it can provide a qualitatively useful guide when combined with a coagulation/fragmentation model and the principle that host patch size controls consumer transmissibility by way of localized Malthusian catastrophes.

The distribution of areas (Figure 3.12) and the perimeter-area relationship (Figure 3.11) together provide an approximation to the complexity profile of the host-patch system, as we defined it back in Chapter 2. Crudely speaking, the interesting part of a host patch is its boundary: in order to say what the patch might do next, we need to know what is happening at its edges. Therefore, the effective information content of a host patch should scale, roughly, with its perimeter. If we neglect the influences of one patch upon another, we can approximate the host population as a collection of blocks, and we can invoke the sum rule stated in §2.1.2. Each block contributes a rectangle to the complexity profile. The width of the rectangle is the number of hosts in the patch, and the height is the information content, which is given by the perimeter. Alternatively, by interchanging the axes, we can construct an MUI curve in the same way. In either case, the structure index so defined will only be an approximation.

3.4 Discussion

But the problem, here, is that it's a form of adaptation that hasn't been studied enough in animals and plants, which is that each change in the species changes what we call the environment, so there is a co-evolution of organism and environment. [...] The organism by its evolution changes the conditions of its life and changes what surrounds it. Organisms are always creating their own hole in the world, their own niche.

—Richard Lewontin [140]

Understanding the effects of spatial extent is a vital part of evolutionary ecology. Spatial extent changes the quantitative and qualitative characteristics of a model's

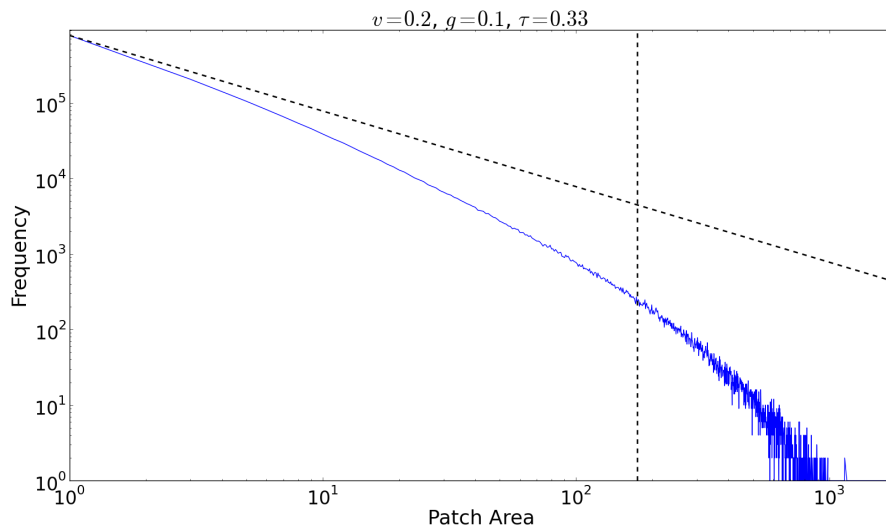


Figure 3.12: Number of host patches observed as a function of patch area ($g = 0.1$, $v = 0.2$, $\tau = 0.33$). The vertical dashed line indicates the 99th percentile, that is, the point at which 99% of the patches are that size or smaller. The sloping dashed line indicates a decay with the inverse of area, for comparison purposes.

evolutionary behavior, compared to well-mixed models. The short-term success rate of novel genetic varieties is not indicative of their long-term chance of success relative to the prevalent type. Standard stability criteria fail to reflect the actual stability achieved over time. We must instead consider extended timescales because they are determined by spatial patterns, whose ongoing formation is an intrinsic part of nonequilibrium evolutionary dynamics. Our analysis provides a clear understanding of why there are dramatic differences between spatial models and mean-field models, which simplify away heterogeneity through mixing populations, averaging over variations or mandating a globally connected patch structure. We have further shown that transplanting organisms dramatically changes the dynamics of spatial systems, even when we preserve local correlations as would be considered in a pair approximation treatment. Our results prove that any model striving to capture the effects of heterogeneity that does not change its behavior with organism transplanting cannot fully capture the dynamics of spatial evolution. The following subsections summarize the general conclusions we draw from these results.

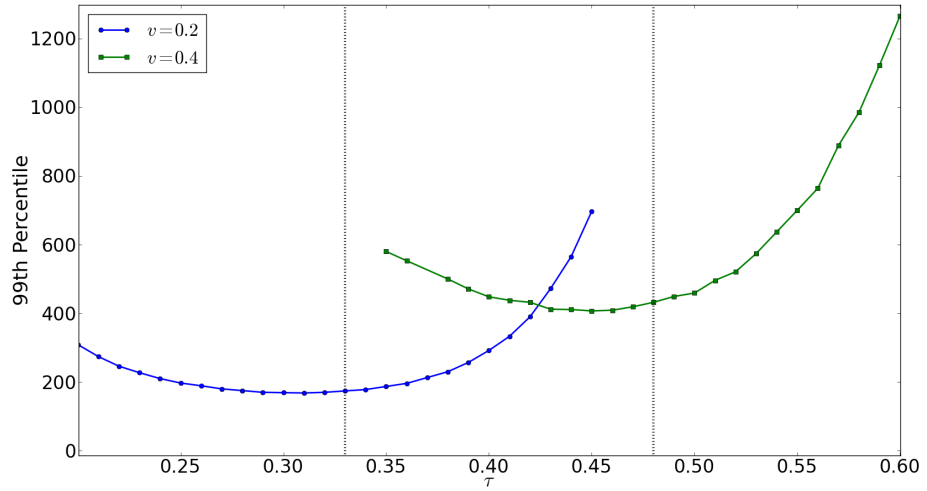


Figure 3.13: 99th percentile of host patch area versus consumer transmissibility τ , for $v = 0.2$ and $v = 0.4$ (computed with $g = 0.1$). The dashed vertical lines indicate the average transmissibilities which evolve when v and g are fixed.

3.4.1 Defining Fitness

In our host–consumer models, each individual either survives or it does not, and any individual has a specific number of offspring and survives over a certain amount of time; that is to say, an “individual fitness” (in the terminology of [141]) is a well-defined concept. To find *expected* individual fitness, or *average* individual fitness, we must define a set of individual organisms over which to take an average, which is the very concept we have established to be problematic. Consequently, derived notions of fitness, which depend on comparisons between such averages [141], become elusive, context-dependent quantities. The problem is both temporal and spatial: Average relative fitness in one generation is not necessarily a good measure of the long-term success of a strain in one, or a combination of, the broad variety of dynamically-generated niches. This problem is not, however, the same as the traditional concept of variation of fitness across a static set of niches, because the niche dynamics of our spatially explicit model ensures that evolutionary outcomes are not reflected in any standard definition of the average.

In the previous chapter, we examined the idea of “frequency-dependent fitness” [44, 45], and one might be inclined to apply that term to voracious invasive strains in this system, as the invasive strain is successful initially when rare but fails when it becomes more common. The term “frequency-dependent fitness” is, however, a misnomer in this context, because the organism type is rare and successful when it is

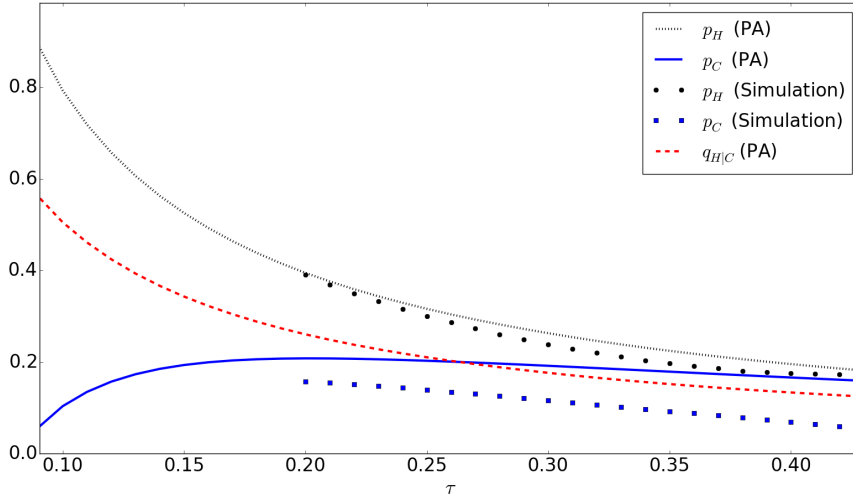


Figure 3.14: Host and consumer densities as a function of τ , with $g = 0.1$ and $v = 0.2$. Note that the agreement between the simulation results and the pair approximation is better for the host density than for the consumers. The pair approximation for this ecosystem was developed by de Aguiar *et al.* [95, 100, 101] and will be discussed in more detail in Chapter 8.

newly introduced, but as it declines to extinction it becomes rare and unsuccessful. Nor can we attribute the decline to the frequency of hosts: the average population density of hosts remains essentially unchanged, because the boom and the following bust are localized. *Frequency*, being defined by an average over the whole ecosystem, is only a proper variable to use for describing the ecosystem in the panmictic case. One might attempt to refine the concept of global frequency by including local frequencies. However, the breakdown of moment-closure techniques implies that defining fitness as a function of organism type together with average local environment [142] will, in many circumstances, not be an adequate solution.

Consequently, we find that trying to assign a meaningful invasion fitness value to an invasive variety of organism is too drastic a simplification. In turn, this implies that we cannot assign a fitness value to a phenotypic or genetic characteristic such as infectiousness or transmissibility. To understand this point, we rephrase the spatial host–consumer model in terms of alleles. In an invasion scenario, an individual consumer can have one of two possible alleles of the “transmissibility gene”, one coding for the native value of τ (*e.g.*, $\tau = 0.33$) and the other for the invasive value (*e.g.*, $\tau = 0.45$). A mean-field treatment would then involve specifying the fraction of the population which carries the native allele versus the fraction which carries the invasive variant. We have seen, however, that the predictions based on such a heavily coarse-

3 Host–Consumer Evolution by Simulation

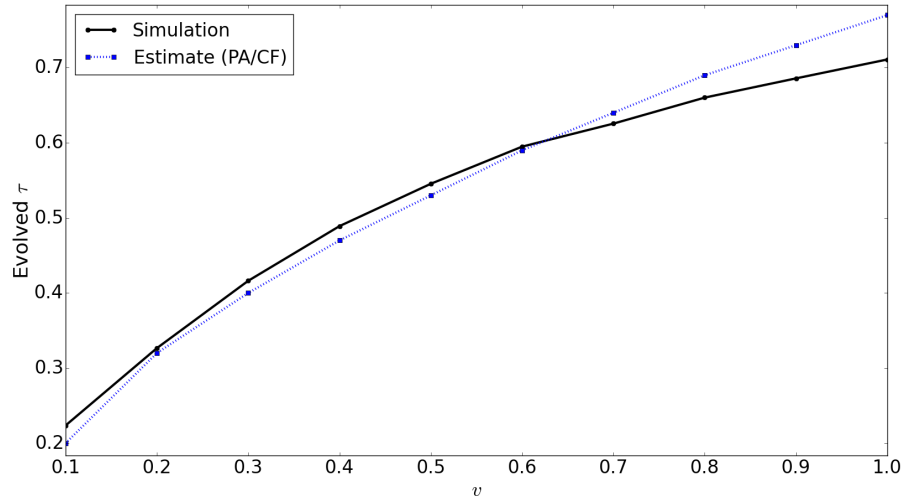


Figure 3.15: Estimates of the evolved transmissibility as a function of the host death rate v (with $g = 0.1$). These values were obtained by combining the results of pair approximation with the insights from the Gueron–Levin coagulation and fragmentation model and the idea that the host–patch-size cutoff controls the evolution of τ . By searching in the pair approximation for a region where $dq_{H|C}/d\tau < 0$ and $dp_H/d\tau > -C$, we identified that part of the τ axis where patch splitting is suppressed and the cutoff elevated. We set $C = 0.75$ based on the $p_H(\tau)$ curves computed for $v = 0.2$ and $v = 0.4$. This value, in turn, provides a reasonable estimate for the evolved transmissibility obtained at other values of v .

grained caricature of the original model deviate from its actual behavior. In short, the evolutionary dynamics cannot be characterized using the allele frequencies at a particular time.

If we can no longer summarize the genetic character of a population by an allele frequency—or a set of allele frequencies for well-defined local subpopulations—then computing the fitness of a genotype from its generation-to-generation change in frequency is a fruitless task. In a world which exhibits nonequilibrium spatial pattern formation, allele frequencies are the wrong attribute for understanding the dynamics of natural selection. Formally, the conventional assumption that the allele frequencies are a sufficient set of variables to describe evolutionary dynamics is incorrect. The spatial structure itself is a necessary part of the system description at a particular time in order to determine the subsequent generation outcomes, even in an average sense.

The timescale-dependence issues which arise in spatial host–consumer ecosystems

exist in a wider context. Multiple examples indicate that initial success and eventual fixation are only two extremes of a continuum which must be understood in its entirety to grasp the stability of a system. In the study of genetic drift, it has been found that neutral mutations can fixate and beneficial mutations fail to fixate due to stochasticity [44]. Likewise, in the study of clonal interference [143, 144], one beneficial mutation can out-compete another and prevent its fixation. Closer to the theme of this chapter, recent work has also emphasized that selection acting at multiple timescales is important for the evolution of multicellularity [145, 146].

Furthermore, classical genetics makes much use of the Price equation for studying the change in a population’s genetic composition over time [46, 47], and it is well known that analytic models built using the Price equation lack “dynamic sufficiency”. That is, the equation requires more information about the current generation than it produces about the next [33, 45, 47, 147, 148], and so predictions for many-generation phenomena must be made carefully, if they can be made at all. Modeling approaches which are fundamentally grounded in the Price equation, such as “neighbor-modulated” fitness calculations [46, 59, 89, 149, 150, 151] and their “multi-level” counterparts [46, 55, 149, 152, 153, 154, 155], are not likely to work well here, as the analyses in question draw conclusions only from the short-timescale regime. In addition, those particular analyses which address host–consumer-like dynamics either rely on moment closures [89] or they assume a fixed, complete connection topology of local populations which are internally well-mixed [59, 152]. These simplified population structures are quite unlike the dynamical patch formation seen in the host–consumer lattice model. (Wild and Taylor [156] demonstrate an equivalence between stability criteria defined via immediate gains, or “reproductive fitness”, and criteria defined using fixation probability; however, their proofs are explicitly formulated for the case of a well-mixed population of constant size, neither assumption being applicable here. Whether fixation probability is equivalent to any other criterion of evolutionary success generally depends on mutation rates, even in panmixia [33].)

We will discuss invasion fitness and moment-closure calculations in more mathematical detail in Chapter 8. The Price equation and the ideas which cluster around it will be our subject in Chapter 9.

In the adaptive dynamics literature, models have been studied in which “the resident strikes back” [157, 158, 159]. That is, an initially rare mutant variety M can invade a resident population of type R , but M does not supplant R and become the new resident variety, even though a population full of type M is robust against incursions by type R . This is often considered a rare occurrence, requiring special conditions to obtain [158, 159], though the theorems proved to that effect apply to nonspatial models, and in adaptive dynamics, it is standard to consider small differences between mutant and resident trait values. The spatial host–consumer ecosystem has the important property that, if mutation is an ongoing process, the spatial extent allows genetic diversity to grow. We initialize the system with all the consumers having the same trait value, but soon enough, different local subpopulations have different trait values. If the effects of single mutations are small, then the different varieties arising have roughly comparable survival probabilities, and so the distribution of extant trait values can spread out. However, the cumulative effect of many mutations which happen to

3 Host–Consumer Evolution by Simulation

act in the same direction on a trait such as transmissibility creates a variety which may engender its own local Malthusian catastrophe. So, the results of rare, big mutations tell us about the spread of trait values we see in the case of frequent, small mutations.

Chapter 6 will treat adaptive dynamics in greater depth, exploring the simplifications which its approximations allow. In essence, adaptive dynamics illustrates the maxim that when one can apply a Taylor expansion, one can simplify. One finds that reductions in complexity depend upon assumptions of smoothness that are remarkably convenient, but are not always applicable.

In our model, transmissibility and consumer death rate are independently adjustable parameters. One can also build a model in which one of these quantities is tied to the other, for example by imposing a tradeoff between transmissibility and virulence of a disease. Different functional forms of such a relationship are appropriate for modeling different ecosystems: host/pathogen, prey/predator, sexual/parthenogenetic and so forth. As long as spatial pattern formation occurs and organism type impacts on the environment of descendants via ecosystem engineering, the shortcomings of mean-field theory are relevant, as are limitations of pair approximations [90].

3.4.2 Pair Approximations and Stability

Pair approximations have been used to test for the existence of an *Evolutionary Stable Strategy* (ESS) in a system—that is, a strategy which, when established, cannot be successfully replaced by another [113]. In addition to the limitations of pair approximation for representing patch structure [104], as we saw in the previous section, the question of whether a mutant strain can initially grow is distinct from the question of whether that strain achieves fixation or goes extinct [81, 92, 115, 143, 160, 161, 162, 163, 164]. The former is a question about short-term behavior, and the latter concerns effects apparent at longer timescales. This distinction is often lost or obscured in analytical treatments. The reason is that one typically tests whether a new type can invade by linearizing the corresponding differential equations at a point where its density is negligible. However, this only reveals the initial growth rate (see the fixed-point eigenvalue analysis in [51, 95, 103, 104]).

Our analysis implies that pair approximations are inadequate for analysis of systems with spatial inhomogeneity. Even including including triple and other higher-order corrections does not suffice, as this series approximation is poorly behaved at phase transitions [111]. Such higher order terms continue to reflect only the local structure of the system and not the existence of well separated areas that diverge in their genetic composition. Nonequilibrium pattern formation will necessarily also be poorly described, at least until the order of expansion reaches the characteristic number of elements in a patch, or an area that encompasses any relevant heterogeneity. Given the algebraic intricacy of higher-order corrections to pair approximations [95, 117, 165], it is useful to know in advance whether such elaborations have a chance at success. As approximation techniques based on successively refining mean-field treatments are blind to important phenomena, we therefore need to build our analytical work on a different conceptual foundation.

3.4.3 Percolation

The mathematical connection between pathogen–host and percolation problems can provide insight into the difficulties in analytical treatment of the biological problem. Spatial heterogeneity gives rise to failure of traditional analytic treatments of percolation and a need for new methodologies. Since the pathogen problem maps onto the percolation problem under some circumstances, the same analytic problems must arise in the biological context. While the presence of a nonequilibrium transition point indicates that traditional analysis techniques fail, it raises the possibility that new tools from the theory of phase transitions [74, 111, 121, 166] will become applicable. For example, in Section 3.3.6, we saw that percolation theory enables us to make quantitatively accurate predictions of population growth and of the critical parameter values which divide one ecological regime from another. Indeed, specific important problems in public health, such as the growth in number of individuals infected in a pandemic, can be considered directly within the context of percolation. Simulations of propagation on approximations of real world networks may help provide accurate predictions, but the general properties of disease propagation can be understood analytically.

Later, in Chapter 4, we will see that percolation models are also helpful in evolutionary game theory, where they yield quantitative results near phase transitions, as they do for host–consumer models. Chapter 7 will develop the mathematics necessary to understand percolation phase transitions. Pursuing this topic in depth will take us into the subject of statistical field theory.

3.4.4 Adaptive Networks

Our results also have significance in the context of adaptive-network research. This field studies systems in which a network’s wiring pattern and the states of its nodes change in interrelated ways. Prior modeling efforts have considered epidemics on adaptive networks, where the spread of the disease *through* the network changes the connections *of* the network [110, 167, 168, 169, 170, 171, 172, 173]. In such models, if a susceptible node has an infected neighbor, it can break that connection by rewiring to another susceptible node. A key point in the analysis is that the new neighbor is chosen at random from the eligible population. This choice of rewiring scheme is exactly what makes a pair approximation work for that epidemic model, because it eliminates higher-order correlations in the system [110]. (Chapter 8 will develop in detail how this method of rewiring allows us to write differential equations for the model.) In our system, by contrast, hosts can form new connections by reproducing into empty sites, but these contacts can only connect geographically proximate individuals.

The difference we have seen between lattice behavior on one hand and RRG or swapping-enabled behavior on the other emphasizes the need to study the effect of spatial proximity on link rewiring. While the structure-erasing nature of unconstrained rewiring among susceptible hosts has been acknowledged [170, 171], new rewiring rules which reflect spatial and community structure have yet to be systematically investigated. The reason that they have not is naturally related to the need for different analytic approaches. “Myopic” rewiring rules, such as restricting the set of eligible new

partners to the neighbors of a node’s current partners, have on occasion been considered, but in contexts other than epidemiology, like evolutionary game theory [174,175], making the endeavour of exploring such rules in epidemic models all the more worth pursuing.

3.4.5 Conclusions

Fisher [176] introduced modern genetic theory in large part motivated by the need to describe the existence of biodiversity.¹ However, the expressions he described which apply in panmictic populations and mean-field treatments lead to a population genetics that rapidly converges to homogeneous populations. Spatial extents and their violation of the mean-field approximations are a key to biodiversity in nature. Their proper theoretical treatment will be a large step forward for evolutionary biology.

Most laboratory experiments, guided by traditional evolutionary thinking, have used well-mixed populations. The results obtained are consistent with theoretical analysis precisely because the conditions are consistent with those assumptions. Such experiments do not provide insight into the role of spatial extent and the implications for real-world biological populations. A growing number of experiments today are going beyond such conditions and, as is to be expected, are obtaining quite different results [64,84,87,92,118,119,164,178,179,180].

Mean-field models are often helpful as a first step towards understanding the behavior of systems, but we cannot trust them to provide a complete story, and we should not let mean-field thinking furnish all the concepts we use to reason about evolutionary dynamics. Our analysis of transplanting organisms can be considered parallel to real world concerns and manifest effects of invasive species introduced by human activity and the impact of shipping and air transportation on pathogen evolution [81,91]. These are among the well-established examples of situations in which spatial extent influences evolutionary dynamics [32,108,164,179,181,182,183,184,185,186,187]. Identifying specific implications of the issues explored in this chapter for particular biological systems [64,84,87,92,118,119,164,178,179,180,188,189,190,191] requires field and laboratory work, as well as theoretical insight to guide the questions that are being asked.

This chapter is based in part on a paper by myself, Andreas Gros and Yaneer Bar-Yam, originally published in October 2011 and updated at the beginning of 2014 [192]. The research reported in that paper was a project I initiated, building on earlier work by Bar-Yam, de Aguiar, Rauch, Sayama, Werfel and others. I wrote the first version of the basic simulation code; Andi figured out how to distribute the work across multiple computers and implemented the organism-swapping idea, which I had after a conversation with him about something tangentially related. The paper-writing process was a collaboration, in which I produced most of the words and my coauthors provided the Darwinian editorial pressure.

¹His *Genetical Theory of Natural Selection* also includes the interesting remark, “No practical biologist interested in sexual reproduction would be led to work out the detailed consequences experienced by organisms having three or more sexes; yet what else should he do if he wishes to understand why the sexes are, in fact, always two?” Biology replies, “Always two?” [177]

4 A Volunteer's Dilemma

The ability of any organism to survive and produce viable offspring depends on its environment, and that environment consists in significant part other living beings. In Chapter 3, we studied this in an evolutionary system based on ecological competition: success required having food to eat. Now, we turn to evolutionary dynamics defined using *game theory*. Interactions among organisms will be represented by mathematical games, and reproductive success will depend on the numerical payoffs won when individuals play those games together.

We can think of the evolutionary game which will be our focus for this chapter as a realization of the general idea discussed at the end of Chapter 2. This game, which we designate the *Volunteer's Dilemma*, is a scenario in which organisms must act in concert to achieve success in the struggle for life.

4.1 Well-Mixed Populations with Carrying Capacity

We introduce the Volunteer's Dilemma in the context of two species living in the same environment. The ecosystem is well-mixed, so we can describe it by population densities. The total number of individuals is restricted, so that the sum of the population densities is bounded by unity:

$$v + s \leq 1. \quad (4.1)$$

Here, v denotes the population density of Volunteers, and s is the population density of Slackers. The game is played in the following way: a group of $K \geq 3$ agents is assembled by randomly drawing from the population. Volunteers pay a cost, c . If all the agents in the group volunteer, then they each gain a benefit, b . A single agent slacking off deprives all the agents from gaining the benefit. Note that we are here taking the “benefit” and “cost” of strategies as given parameters of the model, rather than as emergent consequences derived from some more fundamental dynamics. This is unlike the way in which the idea of competition is realized in the spatial host–consumer system of Chapter 3; we can think of these two different styles of modeling as complementary.

This game is also known as an n -player *stag hunt* [193, 194, 195]. We will stick with the Volunteer's Dilemma terminology in this chapter in order to avoid ambiguity: the stag hunt game is typically defined as a two-player interaction, and so “ n people play a stag-hunt game” could be interpreted as one person engaging in two-player stag hunts with $n - 1$ others, and then receiving the total payoff from these $n - 1$ separate games.

Reproduction in this environment depends upon the availability of empty space, so the probability of a reproduction event diminishes as the total population density $v + s$ increases. We take a pessimistic view of volunteerism: the cost of being a volunteer

4 A Volunteer's Dilemma

must be paid whether or not reproduction happens. This is an idealization of a case where, for example, volunteering requires additional metabolic products which demand heightened energy expenditure regardless of the availability of space to reproduce into. Consequently, we treat the cost c as augmenting the death rate d , while the benefit b augments the growth rate, which is also modulated by the total population density. For convenience, we define $k = K - 1$ to be the number of *additional* Volunteers which a Volunteer must have in its peer group in order to obtain the benefit. We write the following coupled equations for the time evolution of the ecosystem:

$$\dot{v} = -(d + c)v + \left[g + b \left(\frac{v}{v + s} \right)^k \right] v(1 - v - s), \quad (4.2)$$

$$\dot{s} = -ds + gs(1 - v - s). \quad (4.3)$$

We follow Durrett and Levin [196] in having the cost and benefit parameters directly modify the relevant rates. The parameters b , c , d and g are all positive. The increase in Volunteer reproduction due to the benefit b depends on how likely it is that a pool of k agents will contain only Volunteers. A generalization is possible to a case where the size n of the interacting group is larger than the critical number k of Volunteers required to gain the benefit. This would introduce a term of the form

$$b \sum_{j=k}^n \binom{n}{j} \left(\frac{s}{v + s} \right)^{n-j} \left(\frac{v}{v + s} \right)^j. \quad (4.4)$$

All along the line $v + s = 1$, the phase-space flow is inward, because both $\dot{v}$ and $\dot{s}$ are negative. Equations (4.2) and (4.3) have an obvious fixed point at the origin. If $v = 0$, then Eq. (4.3) reduces to

$$\dot{s} = -ds + gs(1 - s), \quad (4.5)$$

which has a stable equilibrium at

$$s^* = 1 - \frac{d}{g}. \quad (4.6)$$

Likewise, if $s = 0$, then the population is all Volunteers, and

$$\dot{v} = -(d + c)v + (g + b)v(1 - v), \quad (4.7)$$

which has a stable equilibrium located on the v -axis at

$$v^* = 1 - \frac{d + c}{g + b}. \quad (4.8)$$

When we move off the axes and consider the v - s plane, then the all-Slacker fixed point is a stable node, while the all-Volunteer fixed point at $(v^*, 0)$ may or may not be stable. Its stability depends on whether or not there exists another fixed point in the off-axis

4.1 Well-Mixed Populations with Carrying Capacity

region.

If both v and s are nonzero, then setting $\dot{s} = 0$ in Eq. (4.3) shows that

$$(v^{**} + s^{**}) = 1 - \frac{d}{g}. \quad (4.9)$$

Setting $\dot{v} = 0$ in Eq. (4.2) and substituting in this expression for the total population density yields

$$\left(\frac{v^{**}}{v^{**} + s^{**}} \right)^k = \frac{cg}{bd}. \quad (4.10)$$

The left-hand side must be less than or equal to unity, so for this to be valid, we must have $cg \leq bd$, or

$$\frac{b}{c} \geq \frac{g}{d}. \quad (4.11)$$

If this is satisfied, then

$$v^{**} = \left(\frac{cg}{bd} \right)^{1/k} \left(1 - \frac{d}{g} \right). \quad (4.12)$$

The fixed point at (v^{**}, s^{**}) is a saddle point, and if it exists, then the all-Volunteer equilibrium at $(v^*, 0)$ is a stable node. Otherwise, the equilibrium $(v^*, 0)$ is a saddle point, unstable to an off-axis push. The location of the all-Volunteer equilibrium on the v -axis is independent of the group size k ; however, its *basin of attraction* is not. The region within which evolutionary trajectories go to $(v^*, 0)$ depends on the location of the coexistence equilibrium, (v^{**}, s^{**}) , and as we have just seen, both v^{**} and s^{**} depend on k . This is another appearance of the theme we encountered in Chapter 3: the stability or instability of fixed points does not tell us everything we need to know about evolutionary dynamics. Note that there is no assortment of like types in this model: the groups within which individuals interact form at random, without regard to types. Nor can the effects of social behaviors on individual fitness be decomposed into a linear combination of pairwise interactions. The success or failure of the Volunteer type is a genuinely and irreducibly synergistic effect.

We will return to this point and see what it means for the traditional language of population biology in Chapter 9. In that context, we will see that the stability criterion we have here, Eq. (4.11), has both familiar and surprising features. To put the matter briefly, that the rule takes the ratio of a benefit parameter to a cost one is a commonplace attribute [197], but what the rule compares that ratio to is not. Furthermore, that the rule does not depend on any kind of assortment among genetically similar individuals is, from a traditional perspective, a surprising outcome. We can interpret this rule as saying that in an ecosystem containing one Slacker and many Volunteers, if a group forms which contains the Slacker, that group will on the whole perform worse than those comprising only Volunteers. Consequently, the Slacker strategy will be penalized by natural selection, if b/c is sufficiently large.

We now establish the stability conditions for the all-Volunteer equilibrium explicitly, by linearizing the system's dynamics near that fixed point. For brevity, we define the

4 A Volunteer's Dilemma

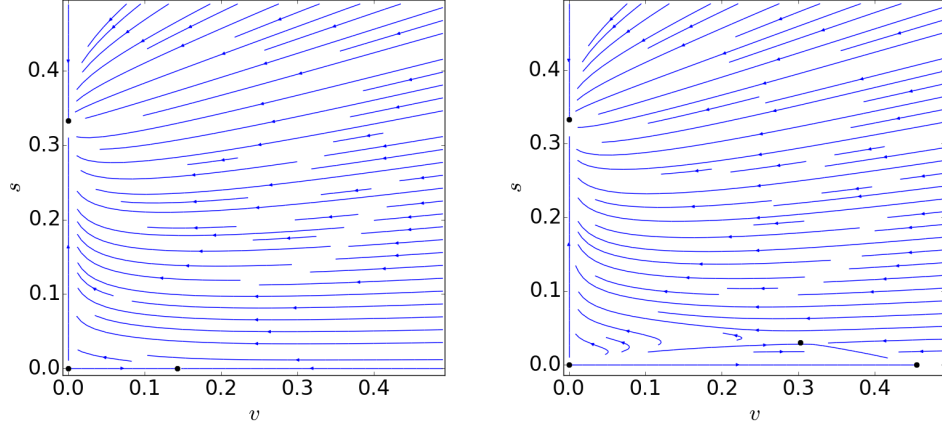


Figure 4.1: **(Left)** Phase portrait for the continuous, well-mixed Volunteer's Dilemma ecosystem, in the regime where Volunteerism is an unstable strategy ($g = 0.15$, $d = 0.1$, $b = c = 0.2$, $k = 3$). Note that of the three fixed points, the origin is unstable, the all-Slacker equilibrium is stable, and the all-Volunteer equilibrium is only stable along the v axis. **(Right)** Phase portrait in the regime where Volunteerism is stable ($g = 0.15$, $d = 0.1$, $b = 0.4$, $c = 0.2$, $k = 3$). Note the presence of an additional fixed point.

abbreviations

$$x = v + s, \quad p = \left(\frac{v}{v + s} \right)^k. \quad (4.13)$$

The Jacobian of the dynamics defined by Eqs. (4.2) and (4.3) is, in terms of these new variables,

$$J = \begin{pmatrix} -(d + c) + \frac{bkps}{x}(1 - x) + (g + bp)(1 - x - v) & -\frac{bkps}{x}(1 - x) - v(g + bp) \\ -gs & -d + g(1 - x - s) \end{pmatrix}. \quad (4.14)$$

At the all-Volunteer equilibrium, all terms proportional to s drop out, and the Jacobian reduces to a conveniently upper-triangular form:

$$J = \begin{pmatrix} (d + c) - (g + b) & (d + c) - (g + b) \\ 0 & -d + \frac{g(d+c)}{g+b} \end{pmatrix}. \quad (4.15)$$

4.1 Well-Mixed Populations with Carrying Capacity

The eigenvalues of this matrix are

$$\lambda_1 = (d + c) - (g + b), \quad (4.16)$$

$$\lambda_2 = -d + \frac{g(d + c)}{g + b}. \quad (4.17)$$

From Eq. (4.8), we know that $d + c$ is always less than $g + b$, if v^* is a valid equilibrium distinct from the origin. Therefore, λ_1 is guaranteed to be negative. The condition for λ_2 to be negative is exactly Eq. (4.11). As promised, the all-Volunteer configuration is a stable one, provided that the ratio of the benefit b to the cost c is greater than g/d , the ratio of the baseline growth and death rates. This rule has the appealing feature that the game-payoff parameters appear on one side, and the baseline rates on the other.

Contrast this with what happens if we define the dynamics using the Prisoner's Dilemma instead. This game, more widely studied than the Volunteer's Dilemma, is a two-player game, rather than a game for an arbitrarily large number of simultaneous players. We recall its definition from Chapter 2: Again, there are two strategies, which we can designate by *Valiant* and *Slinker*. (This terminology is nonstandard, but it lets us keep the same variable names.) Also as before, the interaction payoffs depend upon two parameters, which we can call b and c . Valiant individuals pay a cost c and gain a benefit b if their interaction partner is also Valiant. Meanwhile, Slinkers gain the same benefit from playing with a Valiant, but pay no cost. A Slinker who plays against another Slinker pays nothing and gains nothing. If individuals play with more than one partner during their lifespans, the total benefit they accrue is the sum of the payoffs gained in each instance of the game. We can, therefore, introduce a group-size parameter k , but unlike before, the dependence of growth rates on k will be linear.

As before, we begin by defining a two-variable dynamical system, for which we will then find equilibrium points. Also as before, we let the cost parameter modify the death rate, while the benefit parameter modifies the growth rate. The analogues to Eqs. (4.2) and (4.3) are

$$\dot{v} = -(d + c)v + \left(g + \frac{kbv}{v + s}\right)v(1 - v - s), \quad (4.18)$$

$$\dot{s} = -ds + \left(g + \frac{kbv}{v + s}\right)s(1 - v - s). \quad (4.19)$$

This dynamical system has nonzero fixed points at

$$v = 0, s = 1 - \frac{d}{g}, \quad (4.20)$$

$$s = 0, v = 1 - \frac{d + c}{g + kb}. \quad (4.21)$$

An equilibrium with both v and s nonzero is only possible if c vanishes, which is an uninteresting case.

4 A Volunteer's Dilemma

Can a population of Valiants withstand an intrusion by Slinkers? At the all-Valiant equilibrium, the Jacobian is, by straightforward calculation,

$$J = \begin{pmatrix} (d+c) - (g+kb) & (d+c) - (g+kb) \\ 0 & c \end{pmatrix}. \quad (4.22)$$

The eigenvalues are $(d+c) - (g+kb)$, which is guaranteed to be negative, and c , which is always positive. Therefore, the all-Valiant equilibrium is *never* stable to off-axis perturbations.

Consequently, we see that the socially cooperative strategy is *always* vulnerable to invasion in the Prisoner's Dilemma, whereas in the Volunteer's Dilemma, it can be stable if the costs and benefits of playing the game satisfy a criterion which depends on the death and growth rates. This underlines the importance of studying a variety of games, both two- and multiplayer, in evolutionary game theory. If we had confined our attention to the Prisoner's Dilemma, we would have missed a dynamical system with interesting features.

4.2 Volunteer's Dilemma in a Networked Population

It has been suggested [56, 198] that this kind of nonlinearity can explain the kinds of evolutionary outcomes which are usually taken as requiring “assortment” among genetically similar individuals. Spatial structure is one way to create such assortment, since limited mobility means that geographically proximate organisms are more likely to be genetically related as well. However, there is no fundamental reason why nature should not present us with *both* nonlinearity *and* spatial structure, so it is only natural to see what happens when both occur together.

There are multiple ways to incorporate spatial structure into the evolutionary game dynamics we have defined here. We will address two possibilities in turn. The first method essentially takes the dynamical system defined by Eqs. (4.2) and (4.3) and spreads it over a lattice. We start with an $L \times L$ square lattice, and we specify that each lattice site can be empty (0), occupied by a Volunteer (V) or occupied by a Slacker (S). Our simulation proceeds in discrete time. At each time step, we pick a site at random. If the chosen site is empty, we do nothing. If it contains a Slacker, we pick a neighboring site at random, and if that site is empty, the Slacker can reproduce into it with probability g . Volunteers reproduce likewise, except that their probability of budding into an adjacent empty site depends on how many of their neighbors are also Volunteers. Specifically, if at least two of the neighboring sites contain Volunteers, the baseline reproduction probability g is augmented by an amount b . Next, if the individual we chose is a Slacker, we kill it off with probability d , and if it is a Volunteer, we kill it with probability $d + c$. One generation is defined to have passed when L^2 sites have been sampled.

Figure 4.2 shows what can happen when we implement these stochastic dynamical rules in a simulation. Having constructed the spatial model in this way, we can analyze it following the ideas presented in the previous chapter.

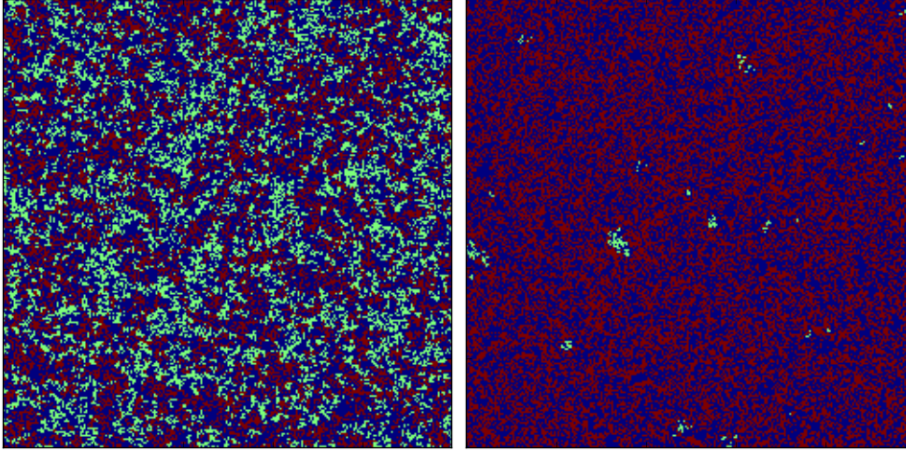


Figure 4.2: **(Left)** Snapshot of the Volunteer's Dilemma simulation on a 250×250 lattice. Red sites indicate Volunteers, cyan sites indicate Slackers and blue sites are empty space. In this example, the death rate is $d = 0.1$, the baseline growth rate is $g = 0.2$, the cost of volunteering is $c = 0.2$ and the benefit is $b = 0.4$. The sites were initialized at random for this illustration and the simulation run for 100 generations. **(Right)** Another snapshot of the Volunteer's Dilemma on a 250×250 lattice, taken 10^3 generations into the simulation. The parameters are the same as at left. Note that Slackers are much less common, but have yet to disappear completely.

Consider a single Slacker placed in an otherwise empty lattice. It will die, leaving an empty space behind, with probability d . If the population is not to vanish, the organism needs to reproduce before then. For budding to be more probable than dying, we must have $g > d$. We expect on general grounds that in practice the critical value of g for having a sustainable population will be somewhat higher than d , thanks to the basic stochasticity of the system. This is known as “fluctuations suppressing the active state.” Simulation results, illustrated in Figure 4.3, bear out this expectation. In each trial, a Slacker population is initialized with a single individual in an otherwise empty world. The number of trials in which the population survives for 500 generations only exceeds 10% for $g > 0.15$. Furthermore, the population density of those trials where the species does survive only grows to significant amounts when g is substantially greater than d .

This pattern repeats for Volunteers: a population grown from an initial seed individual only becomes viable when $g + b$ is significantly larger than $d + c$, as illustrated in Figure 4.3. Again, we see the effect of stochastic fluctuations, which suppress the active phase of the system. Recall that thresholds were also elevated above mean-field expectations in the spatial host–consumer system (Chapter 3, Figures 3.2 and 3.6).

Furthermore, we can understand the behavior near the transition point using per-

4 A Volunteer's Dilemma

colation theory. The most convenient quantity for numerical exploration is $P(t)$, the probability that a population grown beginning with a single seed at time 0 is still alive at time t . We plot a survival probability curve for Slackers in Figure 4.4. The probability fall-off matches that expected for a process which is near criticality in the directed-percolation universality class. Figure 4.5 shows that the same holds true for a population of Volunteers.

We can understand when Volunteerism is favored over Slacking by seeing what happens when both species are present on the same lattice. We begin by filling the lattice at random with 5% Slackers and 5% Volunteers, leaving it otherwise empty. Figure 4.6 indicates that as the benefit parameter b is increased, Volunteers go from unfavored to favored. Note that the crossover occurs when $b/c \approx 1.75$, which is *less* than the crossover point computed in the mean-field model: per Eq. (4.11), the mean-field critical ratio is given by $b/c = g/d = 2.0$. Spatial structure also promotes Volunteerism in another way. In the well-mixed model, an even balance between Slackers and Volunteers will evolve towards the all-Slacker fixed point, even if the all-Volunteer fixed point is stable, unless b is increased still further. That is, points on the line $s = v$ can lie in the all-Slacker equilibrium point's basin of attraction. So, even though the all-Volunteer configuration is stable against invasion, the Volunteer strategy does not predominate when starting from an even balance. However, in the spatial version, this can happen easily.

In addition, as the two strategies become comparable in performance, it takes longer for one to become predominant. As Figure 4.7 illustrates, the time for the losing strategy to vanish from the lattice diverges when b is near the critical transition value. This is reminiscent of the phenomenon known as *critical slowing down* in statistical physics [199, 200]. Examining the crossover region in more detail (Figure 4.8) bears this out. As the performance of the two varieties becomes comparable, the average time required for one to drive the other to extinction increases, and the variation in the time to extinction goes up as well (Figure 4.9). Together, the characteristic exponents and the slowing-down effect demonstrate the relevance of statistical physics to evolutionary game theory.

4.3 Fully Occupied Networks

Another way of incorporating population structure into evolutionary game theory is to put the dynamics on a network and keep each network node occupied. This means that the total population size is constant and equal to the number of nodes in the network. We take up this approach next, as it connects with ways evolutionary game dynamics have been studied before [37, 201].

Suppose that we're trying to figure out the fitness of the i^{th} individual, call it f_i , as a function of the genotypes of the organisms with which it interacts, $\{s_j\}$. To begin with a simplified case, we might make the assumption that the total effect of multiple causes taken together is the sum of the effects those causes would have taken independently, and that the size of the effect grows evenly with the size of the cause.

So, we write a linear equation,

$$f_i = \beta_i^j s_j, \quad (4.23)$$

where we’ve written the parameters with a β as an homage to the “regression coefficient” jargon common to the art. The idea is that if we had a whole bunch of measurements from a laboratory or a field station, we could run a regression analysis and figure out what the values of these coefficients should be. Of course, we can feed any set of numbers we want into the our statistical software package; saying that the results have any predictive value is a stronger statement, which requires making a claim about the linearity of the interactions at work.

An alternative way to measure genotype or trait values is to do so relative to the social circle to which the individual i belongs [46]. This analogous to the physicist’s practice of transforming from the laboratory frame to the centre-of-mass-frame, except we consider instead a “center-of-social-circle” frame. This coordinate change means that the parameters β_i^j get mixed up with each other, but since the equations for the next step of the computations—figuring out how the genetic composition of the population changes as a result of these fitness assignments—are also linear, everything works out pretty simply. We will return to this point in Chapter 9.

A *nonlinear* relationship between genotype and fitness is interesting for both mathematical and biological reasons [47]. One way to see why, which recalls our discussion of Requisite Variety in Chapter 2, is to consider situations where success requires coordinated action. For example, suppose three graduate students are moving across town to a new apartment. They have to transport a heavy object, like a piano or a drill press left in the living room. To move the piano, all three must heft simultaneously and walk in the same direction at the same speed. The payoff to flatmate i is, using $s_i = 0$ to denote “doing nothing” and $s_i = 1$ to indicate “hefting the piano”,

$$f_i = -cs_i + b \prod_j s_j. \quad (4.24)$$

The cost c and benefit b parameterize the situation, which is another realization of the Volunteer’s Dilemma game. As Nowak *et al.* write of the payoff function (4.24), “Clearly in such a game one cannot separate the effects of the action of the second player on the first from the effects of the action of the first on himself” [201]. Nowak *et al.* discuss the three-player version of Eq. (4.24) qualitatively, but they do not treat it in detail. We now take up that analysis.¹

In this section, we will use three different types of connection topology:

- A *regular graph*, in which each node has the same degree k , and the connection pattern is highly symmetric, as in a lattice;
- A *random regular graph*, in which each node has the same degree k , but connections are otherwise unpatterned;

¹They mention the three-player Volunteer’s Dilemma, using the stag-hunt terminology, in an appendix to a paper [201] which provoked quite a bit of controversy [151]. Judging from the sound and fury, the people who got upset about the paper didn’t read the appendices.

4 A Volunteer's Dilemma

- A *mixed graph*, which we generate anew at each time step of the simulation, keeping the node degrees fixed at k .

Simulations were carried out for $k = 3$, $k = 4$ and $k = 6$. The regular graph of degree $k = 4$ was a 20×20 lattice. Both periodic and truncated boundary conditions were studied. For degree $k = 3$, I used a hexagonal lattice of 391 nodes, and to test the effects of boundary conditions, I compared against the results for graph F400A in Foster's census of symmetric graphs [202]. For $k = 6$, a trigonal planar lattice of 400 nodes was constructed, and again both periodic and truncated boundary conditions were studied.

We use a “death-birth” updating scheme [106]. During each generation of the simulation, we progress through the N nodes of the network in random order. At each node, we compute the payoffs of the neighbouring players, using Eq. (4.24). The focal node adopts the strategy of a neighbour chosen stochastically with probabilities based on the neighbours' payoffs. The probability of adopting the strategy of a player who obtained a score f is defined to be proportional to e^{wf} , where the parameter w sets the strength of selection pressure.

Unlike the version of the Volunteer's Dilemma we studied in the previous section, this version allows an individual of one species to displace one of another. Here, no network node is left empty, whereas in the previous version, populations grow into available empty spaces on the lattice.

4.3.1 Initially Balanced Population

The results reported in this section concern what happens when a population is initialized at random with volunteers and slackers (heads it's one, tails it's the other). 2000 trials were performed to obtain each data point, and the number of trials ending in ubiquitous volunteerism was recorded. Error bars in Figures 4.10 and 4.11 indicate two standard deviations.

A few general patterns emerge during the course of numerical simulations:

1. Boundary conditions can promote volunteerism, probably because nodes on an edge or in a corner have fewer neighbours, so it's easier for them to have a neighbourhood full of volunteers.
2. The greater the number of short loops, the larger the difference between the regular and random regular cases.
3. Shuffling the network at each time step always depresses the success rate of volunteerism.

4.3.2 Invasion

We now test *invasion fitness*, by looking at how often a minority strategy can invade a population, rather than how a balanced population goes to one extreme or the other. If neither strategy is favored and the competition is neutral, then any individual in the

population is as likely as any other to become the ancestor of the entire population. Because the future population has to be descended from somewhere, this implies that if an invasive strategy starts with a single individual, the probability it will succeed and sweep to saturation in neutral competition is $1/N$. A strategy which succeeds more often than this is said to be evolutionarily favored [106]. (Other definitions of evolutionary success are possible, and the relationships among them are subtle [33].)

Our model offers two broad categories of invasion scenarios: those in which the resident population is composed of volunteers and those in which it contains only slackers. Overall, we find that the higher the benefit-to-cost ratio b/c , the more favorable the situation for volunteers invading slackers, and the less favorable for slackers invading volunteers. From the results reported above, one would predict a deviation between the outcomes seen on regular lattices from those on RRGs of the same degree if the regular lattices contain short loops. Comparing the results for $k = 3$, Figure 4.12, to those for $k = 6$, Figure 4.13, we find this prediction confirmed. We also see that the b/c ratio necessary to make a volunteer population robust against invasion increases with the network degree k .

4.3.3 Mutation–Selection Equilibrium

Another way to see the effect of closed loops in the underlying graph topology is to compare the steady-state distributions for graphs of the same vertex degrees. If there is a nonzero probability of mutation, then a uniform population does not have to stay uniform: Volunteers can spontaneously appear in a lattice filled with Slackers, and vice versa. (We discussed this for frequency-dependent selection in panmictic populations back in Chapter 2.) Evolutionary success means, in this context, being the more common strategy.

Figure 4.14 demonstrates that short closed loops can make or break an evolutionary strategy. This is a direct indication of the important role which the graph topology plays, and it is also a signal that analytical techniques which are viable for one graph can fail for another.

4.3.4 Analytical Results

It may be possible to deduce some results analytically for the Volunteer’s Dilemma on completely filled graphs, at least in certain limiting cases. Our plan is to follow the logic of earlier work on two-player games [37] and extend it to multiplayer scenarios. The key is to calculate the *expected payoffs* obtained by individuals at various distances from a chosen, focal node. By comparing these quantities at different distances, a criterion for the success of a strategy can be determined. The criterion which obtains will depend, in general, on the update rule in effect.

For the Prisoner’s Dilemma with DB updating, this method indicates that the Valiant strategy can succeed if

$$\frac{b}{c} > k. \quad (4.25)$$

The nonlinearity of the Volunteer’s Dilemma—that is, the presence of higher-scale

4 A Volunteer's Dilemma

structure in its mapping of strategy combinations to payoffs—makes the situation more complicated. The numerical results (Figures 4.12, 4.13 and 4.14) indicate that a success criterion cannot depend on the vertex degrees alone, because two graph structures with the same uniform degree support different evolutionary outcomes. We expect that this will generally be the case for multiplayer games on graphs. The reason why will become clear as we work through the application of the method.

Pick a node in the network, and designate it the focal node. (We assume that the network has sufficient symmetry that the choice of the focal node is arbitrary, as is the case in regular lattices.) Let the state of the focal node be 1, that is, Volunteerism. Next, choose a second node somewhere else in the network. The latter node will also be the site of a Volunteer if it and the focal agent are related by common ancestry without intervening mutations. That is, an agent other than the focus will have the same strategy as the agent at the focal node if they are *Identical By Descent* (IBD). It is also possible for two agents to have the same strategy even if they are not IBD, by accident of fortuitous mutations.

When considering two-player games, it is known [37] that evolutionary success criteria can be derived in the limit of weak selection and low mutation rate, by considering IBD probabilities derived for *neutral drift*. If there is no actual difference in payoffs between the two strategies, then the dynamics are the stochastic copying of labels without preference. Under neutral drift, death-birth updating reduces to the *voter model*, so named because it is an idealization of voters copying their opinions from their neighbors. With a nonzero mutation rate, it becomes a *noisy voter model* [203]. We note that the noisy voter model closely resembles the imitation dynamic we studied in Chapter 2.

In the case of neutral drift and nonzero mutation rate, there exists a unique and well-defined stationary probability distribution over the possible states of the population. We will derive our results using expectation values evaluated with respect to this probability distribution. Let $\langle s_i \rangle$ be the expected value of site i , conditional on the focal site being occupied by a Volunteer. We will use lower indices to label lattice sites, and superscript indices to denote the number of steps taken in a random walk. For example, $\bar{f}_2$ is the expected payoff to the player at site s_2 , while $\bar{f}^{(2)}$ is the expected payoff for a player located at the end of a two-step random walk which starts at the focal site. From these definitions, it follows that

$$\langle s_0 \rangle = \langle s^{(0)} \rangle = 1, \quad (4.26)$$

as well as

$$\bar{f}_0 = \bar{f}^{(0)}. \quad (4.27)$$

“Evolutionary success” can be defined in multiple ways. For example, we could say that the Volunteer strategy is successful if it has more than a 50% share of the population in mutation-selection equilibrium. Or, we could say that Volunteerism is successful if the probability of its fixation when starting with a population comprising one Volunteer and $N - 1$ Slackers is greater than the Slacker strategy’s fixation probability in the reverse scenario. In the low-mutation-rate limit ($u \ll 1$), these conditions

4.3 Fully Occupied Networks

are equivalent [33]. Let b_0 be the probability that the focal individual reproduces, and let d_0 be the probability that it is replaced by a neighbor. These probabilities will depend on the lattice topology, the current configuration of site values, the benefit b , the cost c and the strength of selection w . Then [37] the Volunteer strategy is favored (for weak selection) if

$$\left\langle \frac{\partial(b_0 - d_0)}{\partial w} \right\rangle \Big|_{w=0} > 0. \quad (4.28)$$

For Death-Birth (DB) updating, which we considered in the numerical work reported above, this condition reduces to

$$\bar{f}^{(0)} - \bar{f}^{(2)} > 0. \quad (4.29)$$

To see why, note that d_0 is constant for DB updating, so we only need to evaluate b_0 . Next, we observe that b_0 will depend on the payoffs f_i of the players at the vertices adjacent to site 0. Let $a_{ij} = 1$ if vertices i and j are adjacent, and 0 otherwise. Then

$$\left\langle \frac{\partial b_0}{\partial w} \right\rangle \Big|_{w=0} = \left\langle \frac{\partial}{\partial w} \sum_i \frac{a_{0i} e^{w f_0}}{\sum_j a_{ji} e^{w f_j}} \right\rangle \Big|_{w=0}. \quad (4.30)$$

Eq. (4.29) follows from the quotient rule. When we carry out the derivative with respect to w , we obtain a pair of terms:

$$\left\langle \frac{\partial b_0}{\partial w} \right\rangle \Big|_{w=0} = \left\langle \sum_i \frac{f_0 a_{0i} e^{w f_0} \sum_j a_{ji} e^{w f_j} - a_{0i} e^{w f_0} \sum_j a_{ji} f_j e^{w f_j}}{\left[\sum_j a_{ji} e^{w f_j} \right]^2} \right\rangle \Big|_{w=0}. \quad (4.31)$$

The first term gives us the $\bar{f}^{(0)}$, and the second gives us the $\bar{f}^{(2)}$, establishing the desired relation.

Using Eq. (4.38) for a hexagonal lattice, we have that

$$\bar{f}^{(0)} - \bar{f}^{(2)} = \bar{f}^{(0)} - \frac{1}{3} \bar{f}_0 - \frac{2}{3} \bar{f}_4 \quad (4.32)$$

$$= \frac{2}{3} (\bar{f}_0 - \bar{f}_4). \quad (4.33)$$

This is positive if the expected payoff to the focal individual is greater than that to an individual two steps away.

The expected payoff to the focal player on a hexagonal lattice is

$$\bar{f}_0 = -c \langle s_0 \rangle + b \langle s_0 s_1 s_2 s_3 \rangle. \quad (4.34)$$

Because the focal player is by assumption a Volunteer, this simplifies to

$$\bar{f}_0 = -c + b \langle s_1 s_2 s_3 \rangle. \quad (4.35)$$

4 A Volunteer's Dilemma

What about a player at a distance of one step from the focus? This player will also have three neighbors, one of whom is the focal player. So,

$$\bar{f}_1 = -c \langle s_1 \rangle + b \langle s_1 s_4 s_5 \rangle. \quad (4.36)$$

And a player two steps away from the focus—for example, s_4 —has three neighbors, one of whom is a player who is adjacent to the focus.

$$\bar{f}_4 = -c \langle s_4 \rangle + b \langle s_1 s_4 s_6 s_7 \rangle. \quad (4.37)$$

The site s_1 is one step from the focus, s_4 is two, and s_6 and s_7 are at a distance of three.

A random walk which begins at the focus and has a length of two steps can end either at the focus itself, or at a site two edges away. The first step always goes away from the focus, and the second will return to it with a probability $1/k = 1/3$. By symmetry, we can let s_4 stand in for all the sites at a distance of two steps from the focus. Therefore,

$$\bar{f}^{(2)} = \frac{1}{3} \bar{f}_0 + \frac{2}{3} \bar{f}_4. \quad (4.38)$$

The terms which are moments of a single site variable are easy to evaluate in terms of IBD probabilities. If the focal site and s_i are IBD, then $s_i = 1$. Otherwise, s_i will take the values 0 and 1 with equal probability.

$$\langle s_i \rangle = \bar{q}_i + \frac{1}{2}(1 - \bar{q}_i) = \frac{1 + \bar{q}_i}{2}, \quad (4.39)$$

$$\langle s^{(i)} \rangle = \frac{1 + \bar{q}^{(i)}}{2}. \quad (4.40)$$

Evaluating the cubic and quadratic terms is more complicated. We can make a first stab by assuming they *factor* into products of single-variable expectation values. This is a mean-field approximation. It's difficult to say how drastic its effects will be without doing the calculation, so we press forward.

This is where the fundamental difference with the two-player Prisoner's Dilemma becomes manifest. For that game, we have

$$\bar{f}^{(n)} = -c \langle s^{(n)} \rangle + b \langle s^{(n+1)} \rangle, \quad (4.41)$$

where $\langle s^{(n)} \rangle$ refers to the expectation value of a site which at the end of an n -step random walk starting at the focus. The only expectation values which appear are linear, and we do not have to account for higher-order correlations. We can, therefore, evaluate directly:

$$\bar{f}^{(n)} = \frac{1}{2} \left(-c + b - c\bar{q}^{(n)} + b\bar{q}^{(n+1)} \right). \quad (4.42)$$

And, in turn, the IBD probabilities $\bar{q}^{(n)}$ can be found for symmetric graphs of arbitrary degree. From this, one can deduce the success criterion, Eq. (4.25).

4.3 Fully Occupied Networks

Our primary concern, however, is the Volunteer's Dilemma on a hexagonal lattice. The IBD probabilities we need are given by

$$\bar{q}^{(1)} = 1 - u(N - 1), \quad (4.43)$$

$$\bar{q}^{(2)} = 1 - uN, \quad (4.44)$$

$$\bar{q}^{(3)} = 1 - uN - u(N/3 - 1). \quad (4.45)$$

These can be found iteratively, through a fairly simple argument. Let $p^{(n)}$ denote the probability that an n -step random walk returns to its starting point. This is one way that a site at the end of a random walk can be IBD to the focus: if that site *is* the focus. Therefore, one contribution to $\bar{q}^{(n)}$ will be $p^{(n)}$. The other way two individuals can be IBD is if one is the offspring, without mutation, of a parent which is IBD to the other. Let s_i be a site, distinct from s_0 , which is separated from s_0 by an n -step random walk. The parent of the individual at s_i is an organism at a site s_j which is separated from s_0 by a walk which has $n + 1$ steps. Furthermore, this walk does not return to its starting point at the n^{th} step. By combining these two ways to be IBD, we arrive at the relation

$$\bar{q}^{(n)} = p^{(n)} + \left(\bar{q}^{(n+1)} - p^{(n)} \bar{q}^{(1)} \right) (1 - u). \quad (4.46)$$

This somewhat heuristic argument can be made precise [37], with the same result.

In the limit $n \rightarrow \infty$, symmetry implies that $p^{(n)} \rightarrow 1/N$, and $\bar{q}^{(n)}$ tends to some value which we write $\bar{q}$. Therefore,

$$\bar{q} = \frac{1}{N} + \left(\bar{q} - \frac{\bar{q}^{(1)}}{N} \right) (1 - u), \quad (4.47)$$

which yields

$$(1 - u)\bar{q}^{(1)} = 1 - Nu\bar{q}. \quad (4.48)$$

This lets us eliminate $\bar{q}^{(1)}$ from Eq. (4.46):

$$\bar{q}^{(n)} = (1 - u)\bar{q}^{(n+1)} + Nu\bar{q}p^{(n)}. \quad (4.49)$$

If mutations do not occur, all individuals are IBD. Sending $u \rightarrow 0$ in an expression for $\bar{q}^{(n)}$ must recover this fact. In the limit of low mutation rate, we consequently have the useful expansion

$$\bar{q}^{(n)} - \bar{q}^{(n+1)} = u \left(Np^{(n)} - 1 \right) + \mathcal{O}(u^2). \quad (4.50)$$

Dropping the higher-order terms and rearranging,

$$\bar{q}^{(n+1)} = \bar{q}^{(n)} - u \left(Np^{(n)} - 1 \right). \quad (4.51)$$

4 A Volunteer's Dilemma

The IBD probabilities we require now follow from the basic relations

$$p^{(0)} = \bar{q}^{(0)} = 1, \quad p^{(1)} = 0, \quad p^{(2)} = \frac{1}{k} = \frac{1}{3}. \quad (4.52)$$

This is enough information to compute $q^{(1)}$, $q^{(2)}$ and $q^{(3)}$ by the recursion relation, Eq. (4.51).

Simplifying the $\bar{f}$ quantities by our mean-field approximation yields

$$\bar{f}^{(0)} = -c + \frac{b}{8} \left(1 + \bar{q}^{(1)}\right)^3, \quad (4.53)$$

$$\bar{f}^{(1)} = -\frac{c}{2} \left(1 + \bar{q}^{(1)}\right) + \frac{b}{8} \left(1 + \bar{q}^{(1)}\right) \left(1 + \bar{q}^{(2)}\right)^2, \quad (4.54)$$

$$\bar{f}^{(2)} = -\frac{c}{2} \left(1 + \bar{q}^{(2)}\right) + \frac{b}{16} \left(1 + \bar{q}^{(1)}\right) \left(1 + \bar{q}^{(2)}\right) \left(1 + \bar{q}^{(3)}\right)^2. \quad (4.55)$$

The next step is to substitute in the IBD probabilities. Then, to see whether Volunteerism can succeed with DB updating, we check the difference $\bar{f}^{(0)} - \bar{f}^{(2)}$. We only care about the results to first order in the mutation rate u , because our recursion relation (4.51) is only valid in that limit.

Calculating $\bar{f}^{(0)} - \bar{f}^{(2)}$ is a matter of straightforward but tedious algebra. The result has a form which makes sense, but as we shall see, the numerical details are problematic. For the hexagonal lattice, to first order in the mutation rate u ,

$$\bar{f}^{(0)} - \bar{f}^{(2)} = Nu \left(\frac{5b}{6} - \frac{c}{2} \right). \quad (4.56)$$

This implies that for DB updating, Volunteerism can succeed if

$$\frac{b}{c} > \frac{3}{5}. \quad (4.57)$$

A criterion in terms of the ratio b/c makes sense: it is much like the result we found in the well-mixed case, back in Eq. (4.11). However, the threshold we have found is less than one, which is a puzzling feature. Can the Volunteer strategy really succeed if adopting it brings less benefit than cost?

This strongly suggests that our mean-field approximation was too severe.

Before moving on, we can make one observation. On the hexagonal lattice, it takes a minimum of three steps for separate paths from the focus to converge. At shorter distances, the hexagonal lattice looks like a Cayley tree graph, where all the vertices have degree $k = 3$. The logic we used to find $\bar{f}^{(0)}$, $\bar{f}^{(1)}$ and $\bar{f}^{(2)}$ in terms of expectation values works just as well on the $k = 3$ Cayley tree graph. However, due to the presence of short closed loops, this will not be the case for the square or triangular lattices. Even without being able to compute the expectation values, we can say that under DB updating, Volunteer success on a hexagonal lattice will resemble that on the $k = 3$ Cayley tree, whereas differences will manifest between the square lattice and the $k = 4$ tree, and likewise for the triangular lattice and its tree counterpart. This is, indeed,

what we observe in Figures 4.10, 4.11, 4.12 and 4.14.

One way to make progress, in the absence of a more sophisticated calculation of higher-order expectation values, is to combine the analytical and numerical techniques in a hybrid approach. We can simulate *the neutral-drift process*, and use the results of these Monte Carlo computations to assign values to $\langle s_4 \rangle$, $\langle s_1 s_2 s_3 \rangle$ and $\langle s_1 s_4 s_6 s_7 \rangle$. Then, we can deduce a threshold benefit-to-cost ratio:

$$\left(\frac{b}{c}\right)_{\text{crit}} = \frac{1 - \langle s_4 \rangle}{\langle s_1 s_2 s_3 \rangle - \langle s_1 s_4 s_6 s_7 \rangle}. \quad (4.58)$$

In this way, we can obtain a result valid for the weak-selection regime, using data from the case of *no* selection.

Carrying out this procedure, we find

$$\left(\frac{b}{c}\right)_{\text{crit}} = 4.05 \pm 0.37. \quad (4.59)$$

This value was obtained by simulating 10,000 generations of the neutral-drift dynamics in equilibrium (mutation rate $u = 0.05$) and averaging over the appropriate time series to compute values for the moments in Eq. (4.58). A total of 1,000 simulation runs were made, and the critical ratio computed for each one. Eq. (4.59) reports the mean and standard deviation of those results. The average is comfortably above unity (and, indeed, so were the ratios found in each trial). This is a clear improvement upon the 3/5 threshold we found by the mean-field argument, and it is also in the range consistent with the numerical results for the actual Volunteer's Dilemma process (Figures 4.10 and 4.12).

Consequently, we can say that the hybrid approach is fairly effective. It also has the advantage that Monte Carlo results for the neutral drift process can be applied to multiple different evolutionary games. When we study different games, the success criterion will depend on the fitness function and the update rule, but expectation values of the same form will appear, and so knowledge of the neutral-drift process will be useful in all cases. This means that we can run Monte Carlo simulations once and apply the results to a variety of scenarios, potentially saving ourselves quite a bit of computer time.

One way to improve the rather naïve mean-field factorization of the expectation values is to introduce two-site correlations. We define the pairwise correlation function

$$\langle s_i s_j \rangle_c = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle. \quad (4.60)$$

Recall that each of the quantities on the right are evaluated with the condition that $s_0 = 1$. Therefore, this condition applies to the correlation function on the left as well. If, given that $s_0 = 1$, the sites s_i and s_j fluctuate completely independently, then $\langle s_i s_j \rangle_c$ vanishes.

There is a specific, principled and well-established way to simplify a higher-order expectation value in the approximation that no correlations more complicated than Eq. (4.60) are significant. We will review this method in Chapter 5. For the moment,

4 A Volunteer's Dilemma

we simply avail ourselves of the results. First, the quantity which governs the expected benefit to the focal individual is

$$\langle s_1 s_2 s_3 \rangle \approx \langle s_1 \rangle \langle s_2 \rangle \langle s_3 \rangle + \langle s_1 \rangle \langle s_2 s_3 \rangle_c + \langle s_2 \rangle \langle s_1 s_3 \rangle_c + \langle s_3 \rangle \langle s_1 s_2 \rangle_c. \quad (4.61)$$

This expression reduces to the mean-field, factorized version if the pairwise correlations vanish. And, for the higher-order expectation value we need in order to compute $f^{(2)}$, we obtain the sum

$$\begin{aligned} \langle s_1 s_4 s_6 s_7 \rangle \approx & \langle s_1 \rangle \langle s_4 \rangle \langle s_6 \rangle \langle s_7 \rangle \\ & + \langle s_1 s_4 \rangle_c \langle s_6 s_7 \rangle_c + \langle s_1 s_6 \rangle_c \langle s_4 s_7 \rangle_c + \langle s_1 s_7 \rangle_c \langle s_4 s_6 \rangle_c \\ & + \langle s_1 s_4 \rangle_c \langle s_6 \rangle \langle s_7 \rangle + \langle s_1 s_6 \rangle_c \langle s_4 \rangle \langle s_7 \rangle + \langle s_1 s_7 \rangle_c \langle s_4 \rangle \langle s_6 \rangle \\ & + \langle s_4 s_6 \rangle_c \langle s_1 \rangle \langle s_7 \rangle + \langle s_4 s_7 \rangle_c \langle s_1 \rangle \langle s_6 \rangle \\ & + \langle s_6 s_7 \rangle_c \langle s_1 \rangle \langle s_4 \rangle. \end{aligned} \quad (4.62)$$

How good is this improved approximation for neutral drift on the hexagonal lattice? Monte Carlo computations, under the same conditions as before, show that this approximation for $\langle s_1 s_2 s_3 \rangle$ is accurate to within 0.3%, while it somewhat overestimates $\langle s_1 s_4 s_6 s_7 \rangle$, by about 8%.

4.3 Fully Occupied Networks

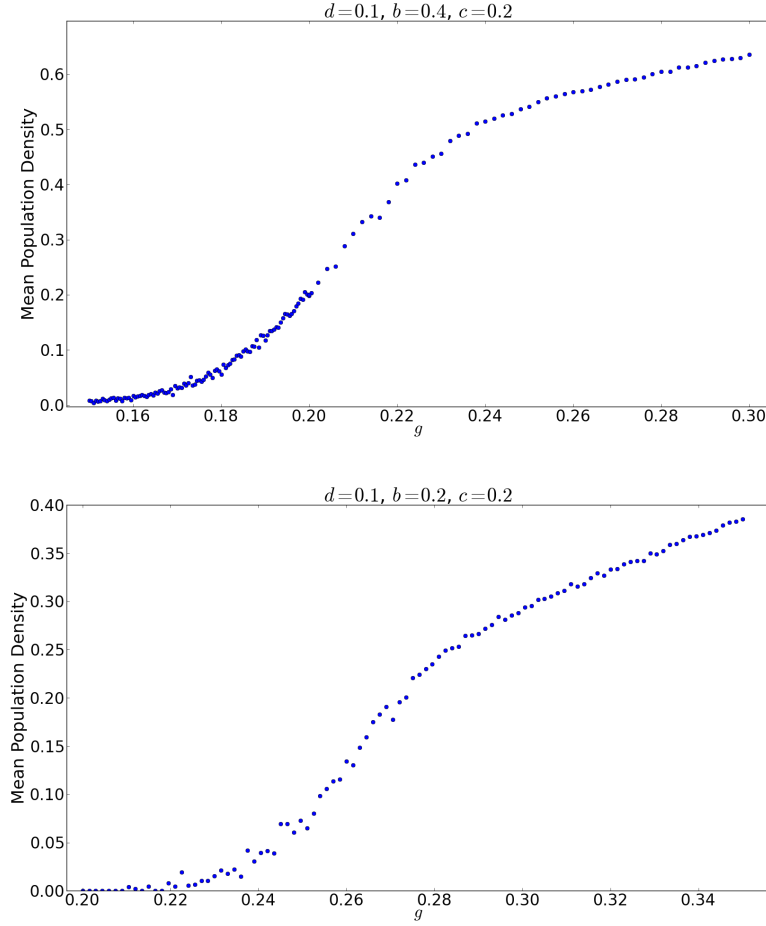


Figure 4.3: **(Top)** Average Slacker density of surviving populations for $d = 0.1$ (and $b = 0.4, c = 0.2$, but those don't matter for Slackers). **(Bottom)** Average Volunteer density of surviving populations for $d = 0.1, b = 0.2$ and $c = 0.2$. In both panels, results were computed with 100 trials per point; each trial was run for 500 generations.

4 A Volunteer's Dilemma

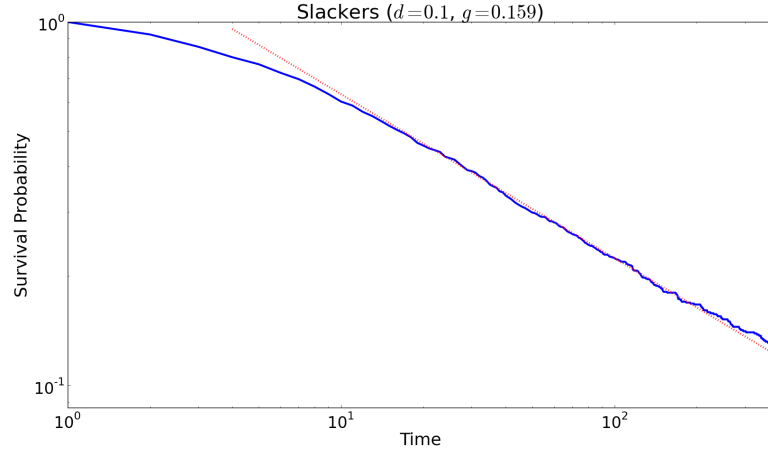


Figure 4.4: Survival probability for a Slacker population initialized with a single seed on a 250×250 lattice ($d = 0.1$, $g = 0.159$). The dotted line shows the characteristic fall-off expected for the directed-percolation universality class, a power-law decay with exponent -0.451 . At longer times, an upward deviation indicates a finite-size effect.

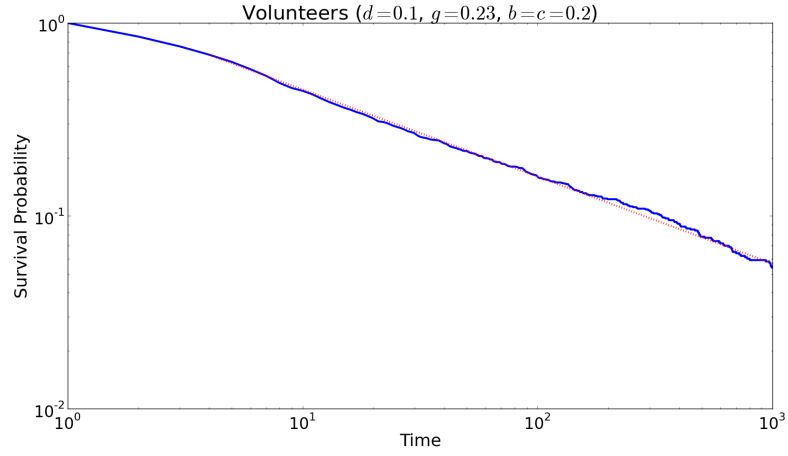


Figure 4.5: Survival probability for a Volunteer population initialized with a single seed on a 250×250 lattice ($d = 0.1$, $g = 0.23$, $b = c = 0.2$). As with the Slacker population in Figure 4.4, the dotted line shows a fall-off characteristic of the directed-percolation universality class, a power-law decay with exponent -0.451 .

4.3 Fully Occupied Networks

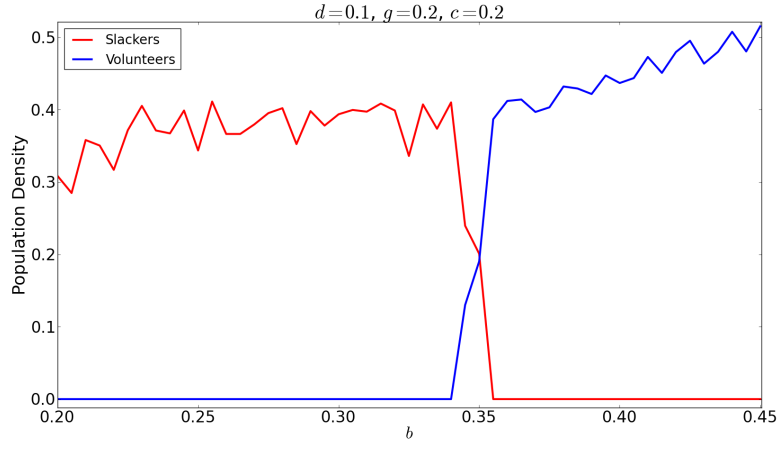


Figure 4.6: Population densities of Volunteers and of Slackers after 10,000 generations or the extinction of the other species, whichever comes first. ($d = 0.1$, $g = 0.2$, $c = 0.2$.)

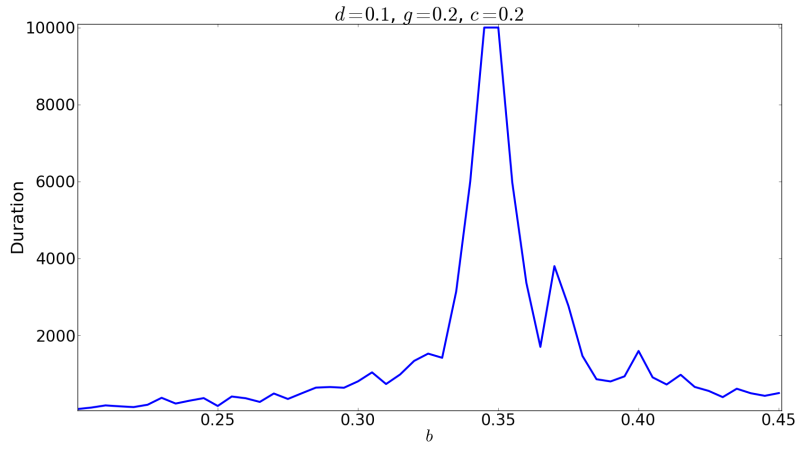


Figure 4.7: Time until simulation completion, set to be 10,000 generations or the extinction of a species, whichever comes first. ($d = 0.1$, $g = 0.2$, $c = 0.2$.)

4 A Volunteer's Dilemma

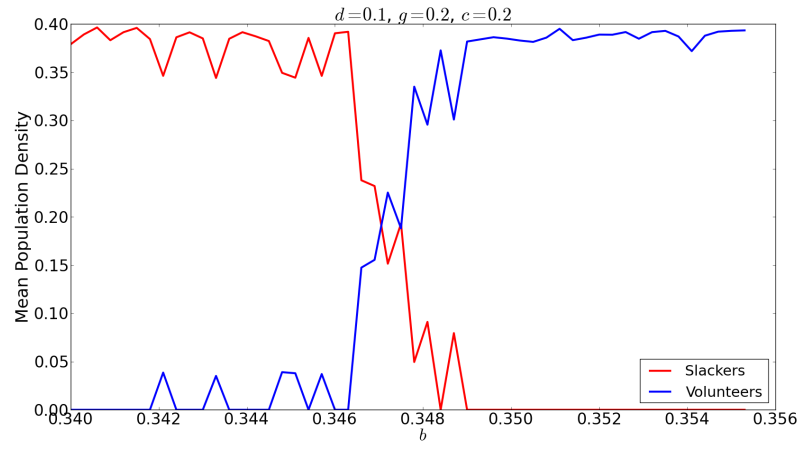


Figure 4.8: Population density of Slackers and of Volunteers, recorded when one of the two populations goes extinct. Computed on a 50×50 lattice, with 10 trials for each value of b . (The other parameters were fixed at $d = 0.1$, $g = 0.2$, $c = 0.2$.)

4.3 Fully Occupied Networks

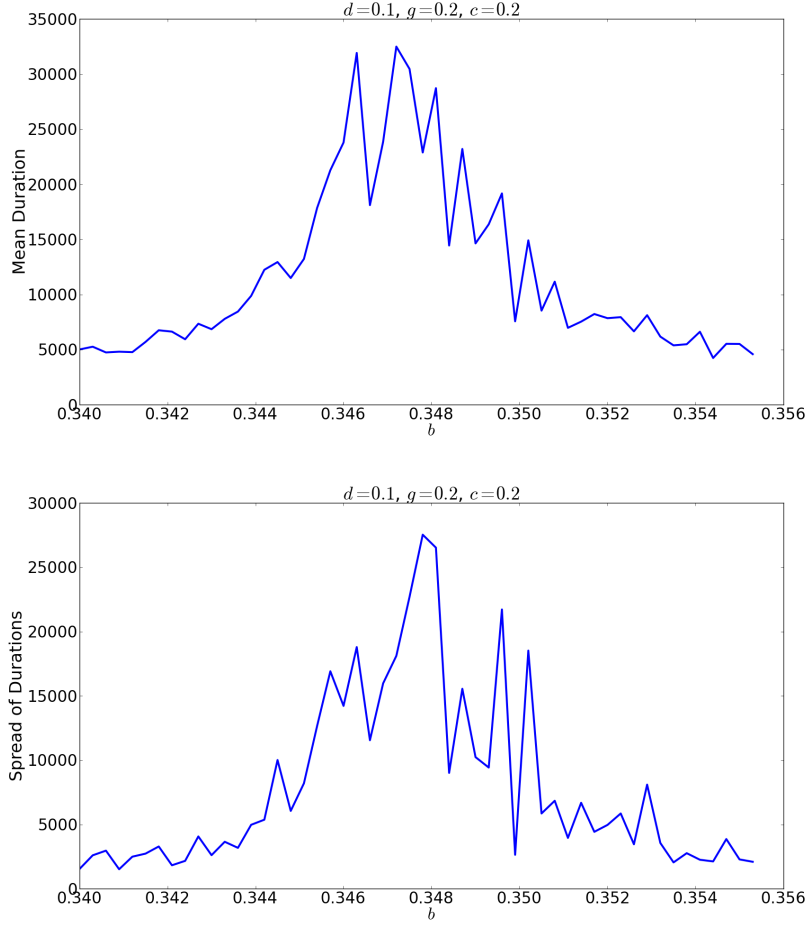


Figure 4.9: **(Top)** Average time for a population to go extinct in the crossover region, signaling the end of a simulation run. **(Bottom)** Standard deviation of simulation durations. Computed on a 50×50 lattice, with 10 trials for each value of b . (The other parameters were fixed at $d = 0.1$, $g = 0.2$, $c = 0.2$.)

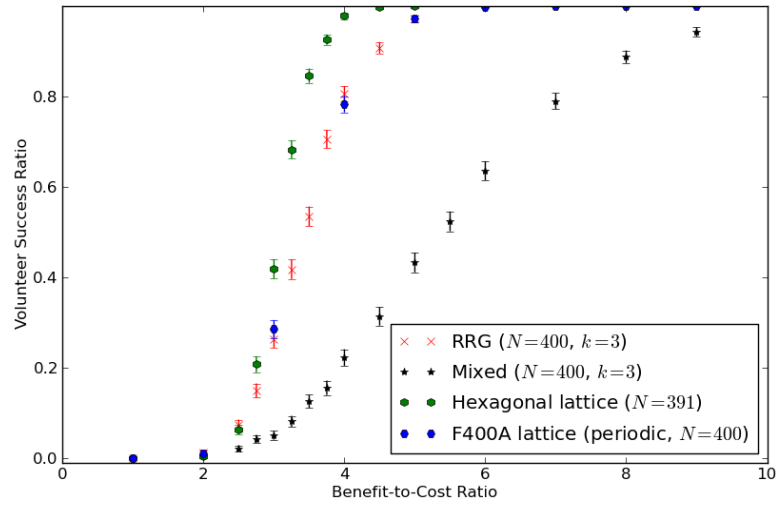


Figure 4.10: Volunteer success rate on various topologies with degree $k = 3$, as a function of benefit b (with cost $c = 1.0$ and strength of selection $w = 0.1$).

4.3 Fully Occupied Networks

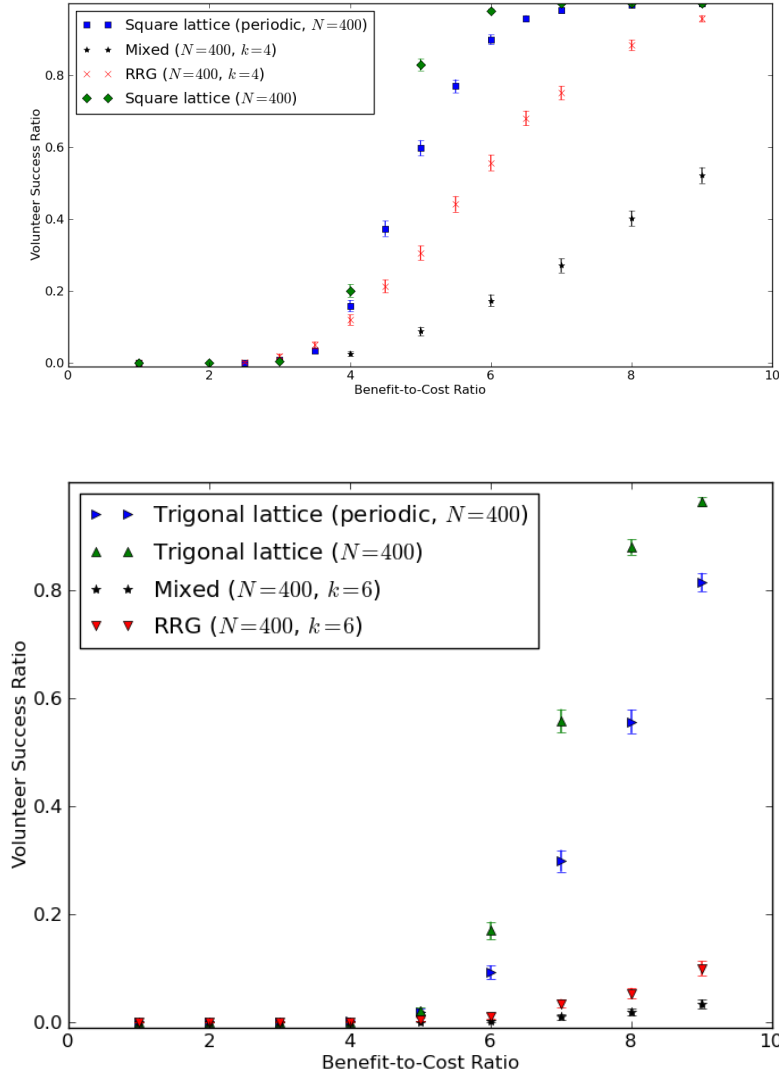


Figure 4.11: **(Top)** Volunteer success rate on various topologies with degree $k = 4$, as a function of benefit b (with cost $c = 1.0$ and strength of selection $w = 0.1$). **(Bottom)** Volunteer success rate on various topologies with degree $k = 6$, as a function of benefit b (with cost $c = 1.0$ and strength of selection $w = 0.1$).

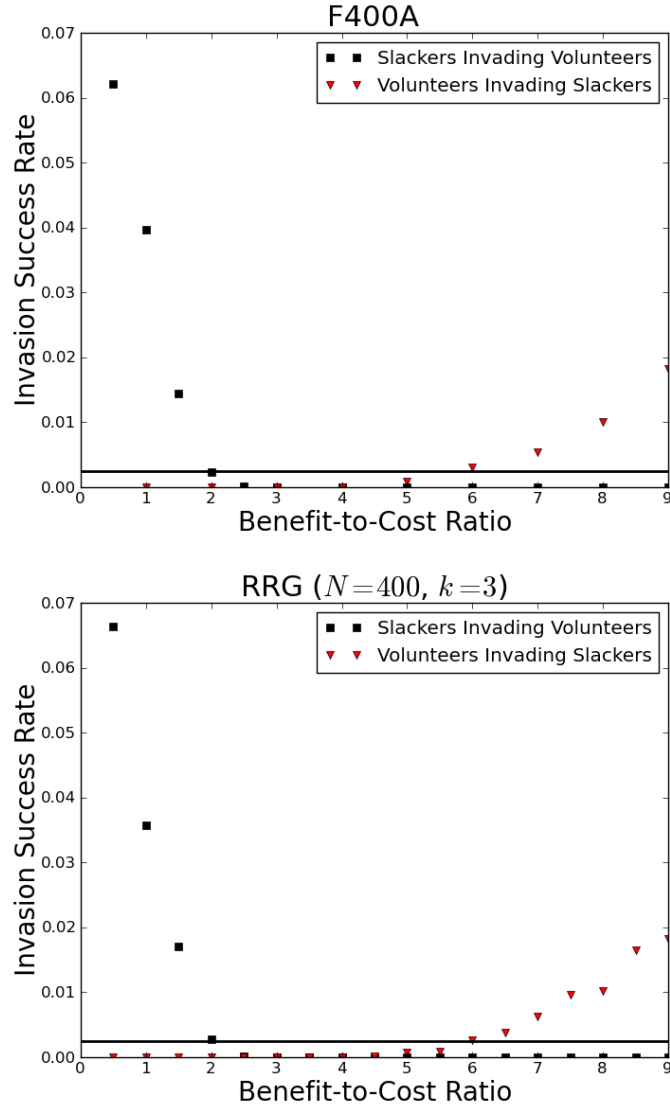


Figure 4.12: **(Top)** Invasion success rate on the F400A topology [202], a regular lattice with degree $k = 3$, as a function of benefit b (with cost $c = 1.0$ and strength of selection $w = 0.1$). The solid horizontal line indicates the expected success rate in the neutral case, $1/N$. **(Bottom)** Invasion success rate on a 400-node RRG with uniform degree 3.

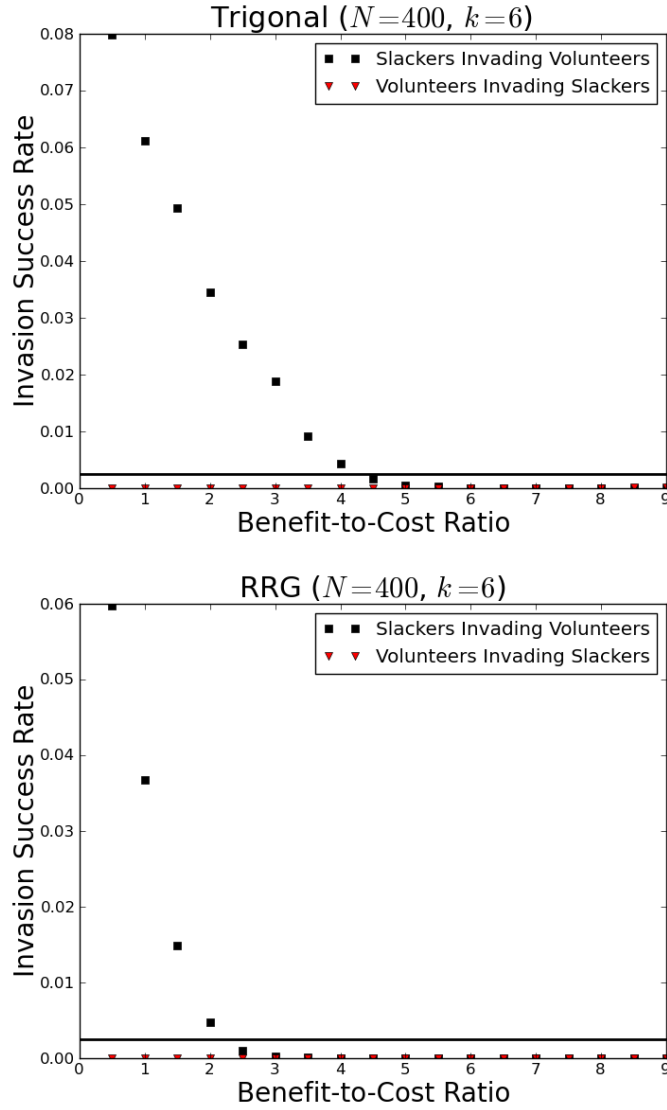


Figure 4.13: **(Top)** Invasion success rate on a 400-node trigonal lattice with periodic boundary conditions, a topology in which all nodes have degree $k = 6$, as a function of benefit b (with cost $c = 1.0$ and strength of selection $w = 0.1$). The solid horizontal line indicates the expected success rate in the neutral case, $1/N$. **(Bottom)** Invasion success rate on a 400-node RRG with uniform degree 6. Note the substantial difference in the slacker-invasion curve as compared to the trigonal lattice case.

4 A Volunteer's Dilemma

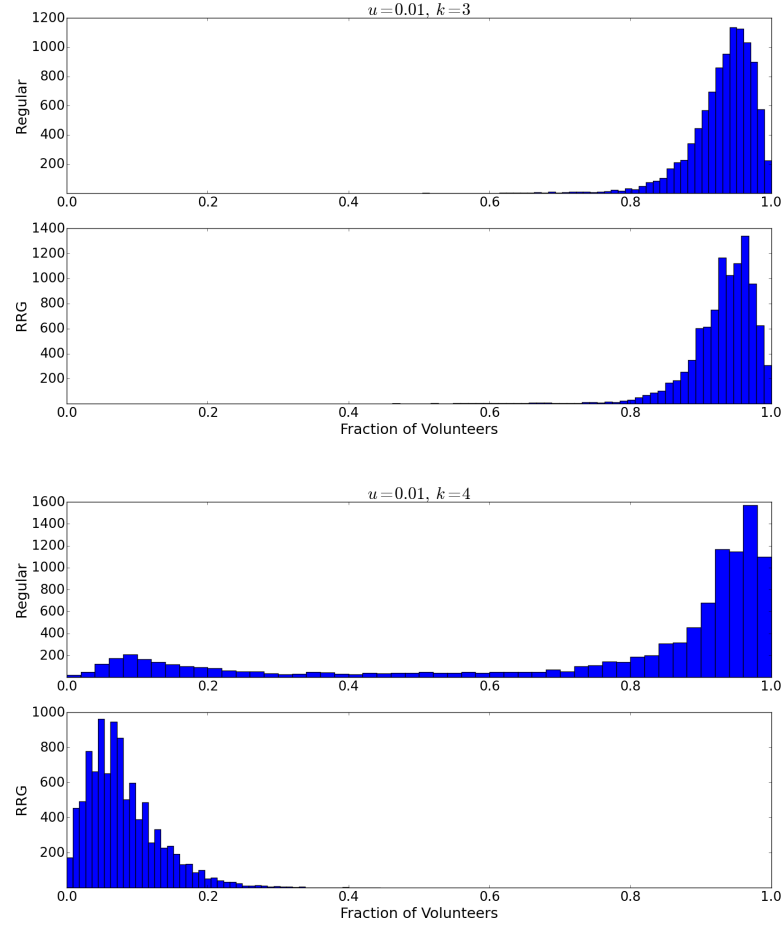


Figure 4.14: **(Top)** Histogram indicating the steady-state distribution of the Volunteer population density, on graphs whose vertices all have degree $k = 3$. (Mutation rate $u = 0.01$, with $b = 4$, $c = 1$, $w = 0.1$; computed for 10,000 Monte Carlo generations.) The distributions for the regular lattice and the random regular graph are quite similar. **(Bottom)** As above, but for graphs with vertex degree $k = 4$ and benefit parameter $b = 5$. Here, on the square grid lattice, the distribution piles up in the majority-Volunteer region, while on the RRG, the population is dominated by Slackers.

5 Techniques of Probability

Probability theory is a way of managing uncertainty and making choices when equipped with incomplete information. In a sense, it is a theory of knowledge and learning, one which is applicable when our attitude about each proposition we consider, or each event we might encounter, can be encapsulated in a real number.

5.1 Basic Properties

One way (but not the only way) to motivate the basic postulates on which this theory relies is a parable: the story of *the android and the Ferengi bartender*. The android walks into a bar, and the bartender, after a quick assessment of the situation, offers a bit of friendly sport between soon-to-be-friends. The bartender proposes to *buy* or *sell* a lottery ticket for any event the android chooses, at any price the android desires. For any event E , the android assigns a numerical weight $p(E)$, which the android deems to be the fair price of the lottery ticket

$$\boxed{\text{Pays \$1 if the event } E \text{ occurs.}} \quad (5.1)$$

The android will *buy* or *sell* this ticket for $\$p(E)$.

The bartender offers the android the opportunity to strike a deal on any combination of events, hoping to catch the android in an inconsistency, forcing a *sure loss*. The goal of the android is to avoid this eventuality.

For example, if the android assigns a price $p(E) < 0$ for some event E , then the android will sell a ticket for a negative amount of money, and the bartender wins. Likewise if the android assigns $p(E) > 1$: this indicates a willingness to sell a ticket for more than it could ever be worth.

Consider two events, E and F , which the android believes are mutually exclusive. The bartender proposes the following three lottery tickets:

$$\boxed{\text{Worth \$1 if } (E \text{ or } F)} \quad (5.2)$$

along with

$$\boxed{\text{Worth \$1 if } E} \quad (5.3)$$

and finally

$$\boxed{\text{Worth \$1 if } F}. \quad (5.4)$$

The value of ticket (5.2) should be the same as the total value of tickets (5.3) and (5.4). If the android professes $p(E \text{ or } F) > p(E) + p(F)$, then the bartender wins. The

5 Techniques of Probability

bartender sells ticket (5.2) and buys tickets (5.3) and (5.4) leaving the android in the red, and whatever happens, the android can never recoup the loss. For example, if event E takes place, the android gains \$1 for the first ticket, but has to pay it back out again by the second.

So, for mutually exclusive events,

$$p(E \text{ or } F) = p(E) + p(F). \quad (5.5)$$

By a similar argument, the android must price the ticket

$$\boxed{\text{Worth } \$\frac{m}{n} \text{ if } E} \quad (5.6)$$

at $\$\frac{m}{n}p(E)$, if m and n are positive integers. A continuity argument shows that a similar statement holds for a positive real number x in place of the rational number m/n .

Furthermore, if “not E ” is the event that E does not happen, then we have two mutually exclusive events, and the sum of the fair prices for their lottery tickets must be the fair price for a ticket that is worth \$1 no matter what happens. Therefore,

$$p(E) + p(\text{not } E) = 1. \quad (5.7)$$

The weights of an event and its complementary event must, to be consistent, sum up to unity.

Now, the bartender proposes a new game: *conditional* lottery tickets. For any two events E and F , a conditional ticket will be worth \$1 if both E and F occur, but the cost of the ticket will be returned if the event E does *not* occur.

$$\boxed{\text{Worth } \$1 \text{ if } (E \text{ and } F). \text{ But money back if } (\text{not } E).} \quad (5.8)$$

This is a gamble on the event “ F given E ,” which we can denote $F|E$. The price of this ticket is, by definition, $\$p(F|E)$. So, this ticket is the same as one written

$$\boxed{\text{Worth } \$1 \text{ if } (E \text{ and } F). \text{ But refund } \$p(F|E) \text{ if } (\text{not } E).} \quad (5.9)$$

To avoid a sure loss at the bartender’s hands, the android must price *this* ticket as the total price of these:

$$\boxed{\text{Worth } \$1 \text{ if } (E \text{ and } F)}, \text{ and } \boxed{\text{Worth } \$p(F|E) \text{ if } (\text{not } E).} \quad (5.10)$$

The price of the final ticket must be $\$p(F|E)p(\text{not } E)$. So, consistency requires that

$$p(F|E) = p(E \text{ and } F) + p(F|E)p(\text{not } E). \quad (5.11)$$

The weights of E and its complement add up to 1, which implies

$$p(F|E) = p(E \text{ and } F) + p(F|E)(1 - p(E)). \quad (5.12)$$

We cancel the $p(F|E)$ from both sides and rearrange to find that

$$p(E \text{ and } F) = p(F|E)p(E). \quad (5.13)$$

We saw earlier that the ticket prices for events deemed mutually exclusive must add. What if the android contemplates two events, E and F , and decides that they are *not* mutually exclusive? In this case, the events “ E and (not F)” and “not F ” are still automatically mutually exclusive in the android’s judgment. Therefore,

$$p([E \text{ and (not } F)] \text{ or } F) = p(E \text{ and (not } F)) + p(F). \quad (5.14)$$

By the distributive rule of Boolean logic,

$$[E \text{ and (not } F)] \text{ or } F = [E \text{ or } F] \text{ and } [(\text{not } F) \text{ or } F], \quad (5.15)$$

which simplifies to

$$[E \text{ and (not } F)] \text{ or } F = [E \text{ or } F]. \quad (5.16)$$

Substituting this into the left-hand side of Eq. (5.14) yields

$$p(E \text{ or } F) = p(E \text{ and (not } F)) + p(F). \quad (5.17)$$

Using the definition of conditional lotteries, we have

$$p(E \text{ and (not } F)) = p(\text{not } F|E)p(E). \quad (5.18)$$

By normalization, this is

$$p(E \text{ and (not } F)) = (1 - p(F|E))p(E). \quad (5.19)$$

Distributing $p(E)$ over the sum and identifying $p(F|E)p(E) = p(F \text{ and } E)$, we find that

$$p(E \text{ and (not } F)) = p(E) - p(F \text{ and } E) = p(E) - p(E \text{ and } F). \quad (5.20)$$

We can substitute this back into Eq. (5.17), yielding

$$p(E \text{ or } F) = p(E) + p(F) - p(E \text{ and } F). \quad (5.21)$$

One way to remember this relationship is in terms of areas: if each event is represented by a geometrical shape, then the total area of the shape representing the event “ E or F ” is the area of the E -shape, plus the area of the F -shape, minus the area of the region where they overlap, which would otherwise be double-counted.

In summary, the requirement that the android avoid a *sure loss* in *any single deal* imposes a rather intricate set of constraints on the fair-price function! Specifically, we have shown that prices must be bounded:

$$0 \leq p(E) \leq 1. \quad (5.22)$$

5 Techniques of Probability

Also, for events believed to be mutually exclusive, the prices add:

$$p(E \text{ or } F) = p(E) + p(F). \quad (5.23)$$

For an event C which the android believes is *certain* to occur,

$$p(C) = 1. \quad (5.24)$$

If the set $\{E_i\}$ is an *exhaustive* set of mutually exclusive events, then

$$\sum_i p(E_i) = 1. \quad (5.25)$$

And *conditional* prices are related to *joint* prices, in accordance with

$$p(F|E) = \frac{p(E \text{ and } F)}{p(E)}. \quad (5.26)$$

It is known [204, 205, 206] that the android can avoid defeat at the hands of the bartender if and only if the price assignment function satisfies Eqs. (5.22), (5.23) and (5.24). An assignment which meets these criteria is called *coherent*.

Note that the bartender does not have to pick any numbers or make any price assignments. All the choices are made by the android; the bartender merely exposes inconsistencies in the numerical weightings which the android attaches to events. Indeed, we could restate the whole scenario as an inner challenge the android poses in the pursuit of self-consistency. Furthermore, note that we have not mentioned *repeated experiments* or *sequences of trials* at all. Avoiding loss at the bartender's hands imposes consistency conditions on numerical weightings, even for events which can only happen once.

The chain of inferences leading to Eqs. (5.22)–(5.24) is typically known as a *Dutch book argument*. The players whom we have designated as the android and the bartender are, in the standard parlance, the bettor and the bookie, and the bettor strives to achieve Dutch-book coherence. The choice of terminology in this section emphasizes the importance of these ideas for machine learning [207, 208, 209], while in addition underscoring the distinction between a mathematicized standard of behavior and the way human beings actually conduct themselves in gambling establishments. Furthermore, the historical origin of the “Dutch book” term is rather obscure, anyway [210, 211].

The coherence conditions (5.22), (5.23) and (5.24) are familiar: together, they say that p is a *probability distribution*. Dutch- or Ferengi-book coherence provides an operational meaning to probability, which as we saw earlier is relevant even for experiments which can be performed only once. We could, on a purely mathematical level, have defined probability axiomatically, declaring that a probability distribution is a set with a particular kind of additional structure. Such a definition [212] might read like the following:

A probability space $(\Omega, \mathcal{B}, \mathbf{P})$ is a tuple comprising a set Ω , a nonempty collection $\mathcal{B}$ of subsets of Ω such that $\mathcal{B}$ satisfies the properties of a σ -

algebra, and a function $\mathbf{P} : \mathcal{B} \rightarrow [0, 1]$ which sends each element of $\mathcal{B}$ to a real number in the unit interval. The function $\mathbf{P}$ is countably additive, and $\mathbf{P}(\Omega) = 1$.

In fact, a mathematician might elect to abstract further from this definition and work with an algebra of random variables and expectation values, eliding the basic set Ω , the event set $\mathcal{B}$ and the mapping $\mathbf{P}$. This turns out to be useful for some subjects, such as random matrix theory [212], but it is not a perspective we will need to pursue for this chapter.

Moreover, we will not have to stress greatly over questions of *continuous* versus *discrete* sets of propositions for events. We will casually switch between discrete probability distributions, normalized as

$$\sum_i p(E_i) = 1, \quad (5.27)$$

and continuous probability densities, which are normalized as

$$\int_{-\infty}^{\infty} dx p(x) = 1. \quad (5.28)$$

The interpretation of the function $p(x)$ is as follows: for any real number x_0 , the quantity $p(x = x_0)dx$ is the probability of the event that x lies between x_0 and $x_0 + dx$. We will not have to be more exacting in our definitions than this. In other words, we will be living by the physicists' standard, which is indifferent to mathematical subtleties until they become unavoidable, treating them as niceties which are no more relevant than the question of how best to construct the real numbers from the integers in the first place [213].

All of our considerations so far in this section have concerned the android's probability ascriptions at *a single time*. Even the conditional probabilities, as in Eq. (5.26), are statements of how the android is willing, at a particular moment, to gamble on events, even if one of those events may chronologically precede the other. We have yet to address how the android might make changes to probability assignments in a self-consistent way.

When we consider probabilities changing with time, our android's probability ascriptions gain a time index. Let p_0 be a function from events to the interval $[0, 1]$ which expresses the android's gambling commitments at the time $t = 0$. Similarly, p_τ denotes the android's gambling commitments at a later time, $t = \tau$. Our problem is to relate p_0 and p_τ according to some standard of consistency.

Some probability assignments may carry two time indices. For example, the bartender can suggest a wager on how many drinks the bar patrons will order in an hour. We can define an event $E(N, T)$ as the event that N drinks will be bought during the interval from time T until one hour later. A probability ascription for this event has one time label in the event itself, and another index which denotes the time at which the ascription is made: $p_t(E(N, T))$. Our concern here is to relate probabilities with different subscripts, which is a problem distinct from the question of how $p_t(E(N, T))$

5 Techniques of Probability

relates to $p_t(E(N, T'))$. Starting in section §5.7, we will address the latter question to a larger extent, and much of Chapter 6 will be devoted to examining it for a particular example.

Suppose that at $t = 0$, the android is willing to gamble on two events, E and D , in such a way that

$$p_0(E|D) = \frac{p_0(E \text{ and } D)}{p_0(D)} = q. \quad (5.29)$$

Now, if the event D occurs between $t = 0$ and $t = \tau$, what should $p_\tau(E)$ be? The simplest choice is to use the number we already have on hand, and force it equal to q :

$$p_\tau(E) = p_0(E|D) = q, \text{ if } D \text{ occurs.} \quad (5.30)$$

However, we *cannot deduce this* from a coherence argument like we have used so far, because all those coherence arguments concern probability assignments *at a single time*. They are, to use a Greek-derived word, *synchronic* statements, when what we need is a *diachronic* rule.

We will now investigate the conditions under which Eq. (5.30) is a reasonable updating scheme. Suppose that at time $t = 0$, the android regards $\$p_0(E)$ as the fair price for a lottery ticket worth \$1 if the event E occurs. Then, at some later time $t = \tau$, the android's evaluation of the world has changed, perhaps in response to new information, and the new fair price is $\$p_\tau(E)$, where $p_\tau(E) < p_0(E)$. There is nothing *irrational* or inconsistent about this: Being willing to sell a ticket at a lower price is just a matter of cutting one's losses.

However, after the android explains this, the bartender proposes a new gamble, this time wagering on the android's own future beliefs. The bartender offers to buy or sell a ticket of the form

$$\boxed{\text{Worth \$1 if } p_\tau(E) = q}. \quad (5.31)$$

Already at $t = 0$, the android can assign a fair price for this ticket, which would be $\$p_0(p_\tau(E) = q)$. If the android is supremely confident that $p_\tau(E)$ will be q , then the fair price of this ticket is \$1.

Now, if $q < p_0(E)$, then the android will be willing to buy a ticket for $\$p_0(E)$ and then sell that ticket for a lower price later. The android faces a sure loss, *one that is already apparent at $t = 0$* , and the bartender has won. Likewise, if $q > p_0(E)$, the android will sell a ticket for $\$p_0(E)$ and buy it back later at a higher price, ensuring a sure loss again.

What can the android do to avoid this eventuality? Defeating the bartender in this scenario requires adhering to the condition

$$p_0(E|p_\tau(E) = q) = q. \quad (5.32)$$

Like all the consistency conditions derived so far, this is a requirement placed upon the current gambling commitments, though in this case, the space of events include the android's future declarations of belief. This condition is an example of the *reflection principle*, an idea due to van Fraassen [206, 214]. It can be derived from a more relaxed

assumption: instead of $p_0(p_\tau(E) = q) = 1$, we can deduce Eq. (5.32) if

$$p_0(p_\tau(E) = q) > 0. \quad (5.33)$$

We are now in a position to apply the reflection principle to the task of choosing a probability-updating scheme. Let E be an event, and let Q_q be the event that at time $t = \tau$, we will have $p_\tau(E) = q$. Next, suppose that there exists a set $\{D_q\}$ of possible *data-acquisition events*, such that there is a bijection between $\{D_q\}$ and the possible values for $p_\tau(E)$. Then

$$p_0(E|D_q) = p_0(E|D_q, Q_q) = p_0(E|Q_q). \quad (5.34)$$

The reflection principle states that

$$p_0(E|Q_q) = p_0(E|p_\tau(E) = q) = q. \quad (5.35)$$

Therefore,

$$p_0(E|D_q) = q. \quad (5.36)$$

This is a statement of a gambling commitment made at time $t = 0$: the android is, at $t = 0$, willing to pay $\$q$ for the lottery ticket

Worth $\$1$ if $(E \text{ and } D_q)$. But money back if $(\text{not } D_q)$.

(5.37)

By definition, q is $p_\tau(E)$ if the event Q_q occurs. Consequently, if the android goes ahead and does what the android had been confident about doing, then

$$p_\tau(E) = p_0(E|D_q). \quad (5.38)$$

This is just the rule we guessed in the first place, Eq. (5.30). It is known as the *Bayes rule*.

We can, again, arrive at this point from a more relaxed assumption. The essential requirement is that the android be able to identify at $t = 0$ an event which, the android expects, can determine the future gambling commitments. That is, there must be some event D such that

$$p_0(p_\tau(E) = q|D) = 1. \quad (5.39)$$

What if no such event D exists, and Eq. (5.39) does not hold? For example, the android may expect that whatever happens between $t = 0$ and $t = \tau$, the data will be too ambiguous to warrant upgrading the confidence in any proposition to 100%. Instead, the android expects that some uncertainty will necessarily remain, which can be represented by a probability distribution over data-acquisition events:

$$\sum_D p_\tau(D) = 1. \quad (5.40)$$

5 Techniques of Probability

In this case, we can use a more general probability-updating scheme, the *Jeffrey rule*:

$$p_\tau(E) = \sum_{D'} p_0(E|D')p_\tau(D'). \quad (5.41)$$

This reduces to the Bayes rule, Eq. (5.30), if

$$p_\tau(D') = \delta_{D,D'}, \quad (5.42)$$

for some event D . And, like Eq. (5.30), the Jeffrey rule (5.41) can be justified using the reflection principle [215].

Diaconis and Zabell [216] give an example where this more general scheme of updating would be applicable:

suppose we are about to hear one of two recordings of Shakespeare on the radio, to be read by either Olivier or Gielgud, but are uncertain as to which, and have a prior with mass $\frac{1}{2}$ on Olivier and $\frac{1}{2}$ on Gielgud. After hearing the recording, one might judge it fairly likely, but by no means certain, to be by Olivier. The change in belief takes place by direct recognition of the voice; all the integration of sensory stimuli has already taken place at a subconscious level. To demand a list of objective vocal features that we condition on in order to affect the change would be a logician's parody of a complex psychological process.

Furthermore [217],

If the *only* impact of hearing the recording is to change the odds on Olivier and Gielgud, in the sense that for any A , $P_\tau(A|O) = P_0(A|O)$ and $P_\tau(A|G) = P_0(A|G)$, then after assessing $P_\tau(O)$ we may proceed to apply Jeffrey's rule. (Of course, the former might well *not* be the case; for example the quality of the recording might convey additional information as to its date or manufacture.)¹

In general, we may say that the updating of probabilities is a subtle subject. The next section will relate this question, which seems to live in the realm of machine learning, to mathematical biology. When we establish this connection, the idea of mesoscale environmental structure will make an appearance.

5.2 An Analogy with Evolutionary Dynamics

Let $\{H_i\}$ be a set of n mutually exclusive and exhaustive hypotheses, so that at any time t ,

$$\sum_{i=1}^n p_t(H_i) = 1. \quad (5.43)$$

¹I have adjusted the notation slightly in this passage to be consistent with the rest of this section.

5.2 An Analogy with Evolutionary Dynamics

In the previous section, we showed that at least under some circumstances, if an event E happens in between times $t = 0$ and $t = \tau$, we are justified in updating the probabilities according to the following rule:

$$p_\tau(H_i) = \frac{p_0(E|H_i)p_0(H_i)}{p_0(E)}. \quad (5.44)$$

This follows from the simple method for updating probabilities, the Bayes rule that we wrote down in Eq. (5.30).

Now, we jump sideways and consider a simple model of evolutionary dynamics in a panmictic population [45]. We suppose there are n types of organism. These could be different species, different genotypes in the same species, or in principle, genetically identical individuals who adhere to different social behaviors. We represent the configuration of the population by an n -tuple of nonnegative real numbers:

$$x = (x_1, x_2, \dots, x_n). \quad (5.45)$$

By assuming panmixia, we deliberately blur over all spatial organization or other kinds of population structure. We neglect stochasticity, and we assume that at each instant, the population configuration is definitely specified. In other words, for this problem we consider evolution as a *deterministic* dynamical system, one which we will take to operate in discrete time. We establish the dynamics by writing an update rule, which yields a new tuple x' when given x .

That a model which neglects all stochasticity should relate to probability theory may be surprising, but we shall soon see that it is the case, and the relationship is quite direct.

To implement the idea of natural selection in this context, we introduce a *fitness function* which maps population tuples to real numbers. Each of the n types has its own fitness function:

$$f_i = f_i(x_1, x_2, \dots, x_n). \quad (5.46)$$

Types which are more fit should be represented more strongly in the next generation. Our update rule for x_i should, consequently, have the form

$$x'_i \propto x_i f_i(x). \quad (5.47)$$

It is convenient to keep the overall population size constant, and in that case, we might as well use the x_i to represent proportions:

$$\sum_{i=1}^n x_i = 1. \quad (5.48)$$

To ensure that this normalization is maintained, we introduce the average fitness

$$\bar{f}(x) = \sum_{i=1}^n x_i f_i(x). \quad (5.49)$$

5 Techniques of Probability

Our dynamical update law, the *replicator equation*, is

$$x'_i = \frac{x_i f_i(x)}{\bar{f}(x)}. \quad (5.50)$$

This is formally analogous [218] to the rule for updating probabilities by conditionalization, Eq. (5.44). To see the relationship, we make the following substitutions:

$$\{p_0(H_i)\} \rightarrow x = (x_1, \dots, x_n), \quad (5.51)$$

$$p_0(E|H_i) \rightarrow f_i(x), \quad (5.52)$$

$$p_0(E) \rightarrow \bar{f}(x), \quad (5.53)$$

$$p_\tau(H_i) \rightarrow x' = (x'_1, \dots, x'_n). \quad (5.54)$$

Natural selection, in the situation modeled by the replicator equation, can be thought of as a learning process, in which the population gains information about the fitness landscape.

What about the more flexible Jeffrey rule? Adapting Eq. (5.41) to this context, we find that

$$p_\tau(H_i) = \sum_{E'} p_0(H_i|E') p_\tau(E') = \sum_{E'} \frac{p_0(H_i) p_0(E'|H_i)}{p_0(E')} p_\tau(E'). \quad (5.55)$$

This reduces to the simpler updating scheme, Eq. (5.44), if $p_\tau(E') = \delta_{E,E'}$. If $p_\tau(E')$ is not a delta function, then multiple potential events E' are relevant. The evolutionary analogue of this is the possibility of *multiple fitness landscapes*.²

Imagine that we have a test tube filled with various types of bacteria. We pipette a fraction w_j of its contents into each of m new test tubes. The conditions in these tubes can differ from one another: perhaps they are illuminated under different frequencies of light, or they contain varying nutrient mixtures. The bacteria interact, and their population proportions change due to natural selection. We then pour the contents of the m tubes all back together again. The total population size remains constant throughout the whole process.

Let $f_i^{(j)}$ be the fitness function for type i in tube j . As before, the fitness depends on the proportions of the different types present in the environment. If all of the populations are scaled down by the same factor w_j , the relative proportions remain the same. Therefore, the initial proportions in each of the new tubes are the same as those in the original source, and we can write the fitnesses as $f_i^{(j)}(x)$. The mean fitness in tube number j is

$$\bar{f}^{(j)}(x) = \sum_i x_i f_i^{(j)}(x), \quad (5.56)$$

and this is true no matter what the value of w_j .

²To my knowledge, the question of an evolutionary analogue of the Jeffrey rule hasn't been raised before. Harper [218] and Baez [219], for example, stop with the Bayes rule, Eq. (5.44).

Because a fraction w_j of each x_i experiences the environment in the j^{th} tube, the new value x'_i will be

$$x'_i = x_i \sum_{j=1}^m \frac{f_i^{(j)}(x)}{f^{(j)}(x)} w_j, \text{ with } \sum_{j=1}^m w_j = 1. \quad (5.57)$$

This is the analogue of the Jeffrey rule, Eq. (5.55).

5.3 Biased Coin-Flips and Urn Models

Consider the prototypical probability problem of repeatedly flipping a coin. Let p be the probability ascribed to the outcome that the coin comes up heads. What probability should we ascribe to the event E_m of seeing the coin land on heads exactly m times out of N ? For this event to transpire, we must see heads m times and tails $N - m$ times. The probability of first seeing m instances of heads, followed by $N - m$ instances of tails, is $p^m(1 - p)^{N-m}$. However, this is not the only way the event E_m can happen: Our specification of the event E_m coarse-grains over the details of the order in which the desired outcomes arrive. So, the probability $p(E_m)$ is the value we computed before, multiplied by the number of ways we can distribute the m heads throughout the total of N flips:

$$p(E_m) = \binom{N}{m} p^m (1 - p)^{N-m}, \quad (5.58)$$

This defines the *binomial distribution*, in which we have used the binomial coefficients [220] more explicitly given as

$$\binom{N}{m} = \frac{N!}{m!(N - m)!}. \quad (5.59)$$

We can think of this scenario in a slightly different way, which leads to some useful modifications. Consider an urn filled with a total of N_B balls, and let $R = N_B p$, where p is the parameter we used before. Exactly R of the balls in the urn are red, the other $G = N_B - R$ being green. We draw a ball from the urn at random, check its color and drop it back into the urn. The probability that the ball we pick will be red is just p . If we repeat the draw-and-replace operation N times, the probability that we see red in exactly m trials is given by Eq. (5.58).

What happens if we do *not* replace a ball after we withdraw it? Then, the population of the urn changes from one trial to the next. Call E_m the event of seeing red in exactly m trials out of N . The probability of this event is

$$p(E_m) = \frac{(\# \text{ of ways to get } m \text{ red balls}) \times (\# \text{ of ways to get } N - m \text{ green balls})}{(\text{total } \# \text{ of ways to make a selection})}. \quad (5.60)$$

Each factor in this expression can be found using combinatorics. In fact, each quantity

5 Techniques of Probability

which enters this formula is a binomial coefficient:

$$\begin{aligned} (\# \text{ of ways to get } m \text{ red balls}) &= \binom{R}{m}, \\ (\# \text{ of ways to get } N - m \text{ green balls}) &= \binom{G}{N - m}, \\ (\text{total } \# \text{ of ways to make a selection}) &= \binom{R + G}{N}. \end{aligned} \quad (5.61)$$

Putting these together, we have

$$p(E_m) = \frac{\binom{R}{m} \binom{G}{N-m}}{\binom{R+G}{N}}. \quad (5.62)$$

This defines the *hypergeometric distribution*. Expanding out the binomial coefficients per Eq. (5.59),

$$p(E_m) = \frac{R!G!}{m!(R-m)!(N-m)!(G-N+m)!} \frac{N!(R+G-N)!}{(R+G)!}. \quad (5.63)$$

The third variation of the urn problem is the following: Every time we draw out a ball, we check its color, and we return *two* balls of that color to the urn. This is the *Pólya urn model* [221, 222], and in this case, if we begin with R red balls and G green balls,

$$p(E_m) = \frac{N!}{m!(N-m)!} \frac{(m+R-1)!(N-m+G-1)!}{(N+R+G-1)!} \frac{(R+G-1)!}{(R-1)!(G-1)!}. \quad (5.64)$$

In terms of the gamma function,

$$p(E_m) = \frac{\Gamma(N+1)}{\Gamma(m+1)\Gamma(N-m+1)} \frac{\Gamma(m+R)\Gamma(N-m+G)}{\Gamma(N+R+G)} \frac{\Gamma(R+G)}{\Gamma(R)\Gamma(G)}. \quad (5.65)$$

This is the beta-binomial distribution we encountered in Chapter 2, as the steady-state solution to imitation dynamics on a complete graph, Eq. (2.34).

These probability distributions are related in another way, other than their common appearance in variations of the urn scenario [223]. We can find this relationship by revisiting the conditionalization rules that we studied in the previous sections. Take the Bayes rule (5.44), which effects a map from one probability distribution to another:

$$p(\theta) \rightarrow \frac{p(y|\theta)p(\theta)}{p(y)}. \quad (5.66)$$

When confronted with a transformation, we typically like to know what it leaves unchanged. For example, the eigenvectors of a matrix are the vectors which the transformation represented by that matrix changes only by a scaling factor. Likewise, it is

5.3 Biased Coin-Flips and Urn Models

useful to identify probability distributions which, under the conditionalization map, keep the same form. The input to the map is known as the *prior*, and its output is the *posterior*. If the prior and posterior have the same functional form, then they are *conjugate*. This relation is dependent on the likelihood function $p(y|\theta)$ used in the mapping, but that can often be taken as fixed on other grounds.

Suppose that

$$p(y|\theta) = \binom{N}{y} \theta^y (1 - \theta)^{N-y}, \quad (5.67)$$

for a positive integer N and all integers y satisfying $0 \leq y \leq N$. Then, the conjugate prior of the binomial distribution (5.58) is the Beta distribution,

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \quad (5.68)$$

and the conjugate prior of the hypergeometric distribution (5.62) is the beta-binomial, Eq. (5.65).

We now have the tools to dig a little more deeply. Consider a scenario in which we wish to *repeat* an experiment. That is, we have the budget to carry out a long experiment, made up of many successive trials [224]. We can represent the outcome of each trial by a random variable x_j , and we can assign a joint probability distribution $p(x_1, x_2, \dots, x_N)$ over the possible outcomes of an N -trial experiment. Motivated by common experiences in the workaday life of a scientist, we now impose two conditions on this joint probability distribution. These conditions will help dramatically in narrowing down the possible forms of the joint distribution $p(x_1, x_2, \dots, x_N)$. First, we require that it be *finitely exchangeable*: its value is invariant under permutations of its arguments. If π is any permutation of the numbers $\{1, \dots, N\}$, then

$$p(x_1, \dots, x_N) = p(x_{\pi(1)}, \dots, x_{\pi(N)}). \quad (5.69)$$

This property recalls the exchange symmetry we invoked in Chapter 2 to simplify the complexity profile.

Second, we require that $p(x_1, \dots, x_N)$ be *extensible*, in the following manner. For any integer $M > 0$, there is a finitely exchangeable distribution with more arguments, p_{N+M} , such that

$$p(x_1, \dots, x_N) = \sum_{x_{N+1}, \dots, x_{N+M}} p_{N+M}(x_1, \dots, x_N, x_{N+1}, \dots, x_{N+M}). \quad (5.70)$$

These two requirements make precise the idea that our probability assignment p derives from an arbitrarily long sequence of random variables, the order of which is inconsequential. We say that a p which satisfies both conditions, finite exchangeability and extensibility, is *exchangeable*.

Let us say we have an exchangeable probability assignment for M binary random variables $x_1, \dots, x_M$. Define $p(n, N)$ to be the probability of obtaining n 1's in N

5 Techniques of Probability

trials. Because the n appearances of the outcome 1 can arrive in any order, this is

$$p(n, N) = \binom{N}{n} p(x_1 = 1, \dots, x_n = 1, x_{n+1} = 0, \dots, x_N = 0). \quad (5.71)$$

This is equal to

$$p(n, N) = \binom{N}{n} \sum_{m=0}^M p(x_1 = 1, \dots, x_n = 1, x_{n+1} = 0, \dots, x_M = 0 | m, M) p(m, M). \quad (5.72)$$

Exchangeability means that all sequences with m occurrences of “1” in M trials are equiprobable. So, we have again an urn problem, specifically, the one that led us to the hypergeometric distribution. If we stock an urn with M balls, m of which are red, then the hypergeometric distribution tells us how likely we are to find n red balls in N draws made without replacement.

$$p(x_1 = 1, \dots, x_n = 1, x_{n+1} = 0, \dots, x_M = 0 | m, M) = \frac{\binom{m}{n} \binom{M-m}{N-n}}{\binom{M}{N}}. \quad (5.73)$$

Defining

$$(r)_q = \prod_{j=0}^{q-1} (r - j) = \frac{r!}{(r - q)!}, \quad (5.74)$$

we have after a spot of algebra that

$$p(n, N) = \binom{N}{n} \sum_{m=0}^M \frac{(m)_n (M - m)_{N-n}}{(M)_N} p(m, M). \quad (5.75)$$

We can rewrite this sum as an integral:

$$p(n, N) = \binom{N}{n} \int_0^1 dz \frac{(zM)_n [(1 - z)M]_{N-n}}{(M)_N} P_M(z), \quad (5.76)$$

where we have defined

$$P_M(z) = \sum_{m=0}^M p(zM, M) \delta(z - m/M). \quad (5.77)$$

Thanks to the extendibility property, we can take the limit $M \rightarrow \infty$. In this limit, we find that $P_M(z)$ approaches a continuous curve, $P_\infty(z)$, and the rest of the integrand goes to $z^n (1 - z)^{N-n}$.

$$p(n, N) = \binom{N}{n} \int_0^1 dz z^n (1 - z)^{N-n} P_\infty(z). \quad (5.78)$$

Interestingly, the result has the form of an integral over what we might call a “meta-probability.” Note that $z^n(1-z)^{N-n}$, times the binomial coefficient out front, has the form of a binomial distribution with probability equal to z . The curve $P_\infty(z)$ has, by construction, the normalization property of a probability density:

$$\int_0^1 dz P_\infty(z) = 1. \quad (5.79)$$

It is as if we can write the function $p(n, N)$ in terms of “the meta-probability $P_\infty(z)$ that the probability is z .”

Eq. (5.78) can be generalized readily enough beyond the case of binary random variables. The result is *de Finetti’s theorem*. Let Δ_k denote the space of valid probability assignments over k outcomes:

$$\Delta_k = \left\{ \vec{p} : p_j \geq 0 \text{ for all } j \text{ and } \sum_j p_j = 1 \right\}. \quad (5.80)$$

Then, exchangeability implies that

$$p(x_1, \dots, x_N) = \int_{\Delta_k} d\vec{p} P(\vec{p}) p_{x_1} \cdots p_{x_N} = \int_{\Delta_k} d\vec{p} P(\vec{p}) p_1^{n_1} \cdots p_k^{n_k}, \quad (5.81)$$

where $P(\vec{p})$ is properly normalized over Δ_k :

$$\int_{\Delta_k} d\vec{p} P(\vec{p}) = 1. \quad (5.82)$$

When we used the parable of the Ferengi bartender to motivate the basic rules of probability, the story left no room for an “unknown probability,” per se. The android’s probability assignments at any time are known to the android, and the idea of an “unknown known” sounds like a contradiction in terms. However, de Finetti’s theorem gives meaning to it. The locution “unknown probability” is a *shorthand for a scenario in which the android is gambling on a sequence of trials that the android judges to be exchangeable*.

5.4 Shannon Information

Intuitively speaking, *information* is that which removes our uncertainty. We can quantify an amount of information by specifying how many questions are necessary to overcome the uncertainty we have about something. For example, suppose we have an experiment which yields one of M different outcomes, stochastically. If we repeat this experiment N times, the number of different possible sequences of outcomes is M to the N^{th} power. Each of these sequences can be represented as a string of symbols, drawn from an alphabet of size M . The number of yes-or-no questions we would need

5 Techniques of Probability

to narrow down this set of possibilities to a single result is $\log_2 M^N = N \log_2 M$. However, we may have *expectations* about the experiment, meaning that we will find some strings of outcomes less surprising than others. Suppose that we ascribe a probability p_i to each of the M possible outcomes of an individual trial in the sequence. We have a set of nonnegative numbers which together satisfy

$$\sum_{i=1}^M p_i = 1. \quad (5.83)$$

This set defines a *random variable*. We do not know in advance exactly how many times the i^{th} outcome will occur, but the *number of such occurrences we will find least surprising* is easily computed:

$$N_i = N p_i. \quad (5.84)$$

How many of the M^N possible strings are, in terms of our probability assignment, unsurprising? A minimally surprising string is one where each of the M symbols in our alphabet occurs N_i times. The number of such strings having total length N is, by elementary combinatorics,

$$K = \frac{N!}{\prod_{i=1}^M N_i!} = \frac{N!}{\prod_{i=1}^M (N p_i)!}. \quad (5.85)$$

How many bits do we need to specify one message out of a set of *this* size? Applying the algebraic properties of logarithms, this is

$$\log_2 K = \log_2 N! - \sum_{i=1}^M \log_2 (N p_i)!. \quad (5.86)$$

Here, it is useful to invoke Stirling's approximation,

$$\log N! \approx N \log N - N. \quad (5.87)$$

With this, we can evaluate the natural log of K as

$$\log K \approx N \log N - N - \sum_i [N p_i \log(N p_i) - N p_i]. \quad (5.88)$$

We have assumed here that each of the N_i is large enough for Stirling's approximation to be viable. Now, we simplify. By normalization, the p_i must sum up to one, so we can cancel the second and final terms, leaving

$$\log K \approx N \log N - N \sum_i p_i [\log N + \log p_i]. \quad (5.89)$$

Invoking normalization again, we see that we are left with

$$\log K \approx -N \sum_i p_i \log p_i, \quad (5.90)$$

which we can easily convert to a base-2 logarithm by division. Then, if we divide by N , we can obtain a result in terms of bits per symbol.

This result provides us with a *measure of information* associated with the probability distribution $\{p_i\}$. If the probabilities p_i express our expectations of the stochastic experiment, then the number of yes-or-no questions which we should expect to require in order to remove our uncertainty about each iteration of that experiment is

$$H[\{p_i\}] = - \sum_i p_i \log_2 p_i. \quad (5.91)$$

This is the *Shannon information* of the probability distribution, also known variously as the Shannon entropy and the Shannon index. The base of the logarithm is a convention which depends on the subfield of science that one is currently studying. The quantity $H[\{p_i\}]$ vanishes if $p_i = \delta_{ij}$ for some j , and it is maximized by the uniform probability distribution over all i . This expresses the fact that if our expectations favor all possibilities equally, we cannot expect to gain any advantage by asking questions cleverly (as we could, for example, with English text: “Is the next letter an *E*?”). Contrariwise, if we are supremely confident that a particular outcome will obtain, we expect that we will require zero questions to ascertain the result.

Back in Chapter 2, we used a general idea of an information function to quantify the concept of multiscale structure. We can apply Shannon information in that context if we consider not just one random variable, but combinations of them. If we have two random variables X and Y and a joint probability distribution $p(x, y)$, the total joint information of X and Y considered together is, by the Shannon formula,

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y). \quad (5.92)$$

If we only care about one of the two random variables, we can sum over the other:

$$p_X(x) = \sum_y p(x, y), \quad p_Y(y) = \sum_x p(x, y). \quad (5.93)$$

In turn, we can use these probability distributions in the Shannon formula just as well.

When we developed our formalism in Chapter 2, we considered a measure of *shared information* which satisfied the relation

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \quad (5.94)$$

Is there a formula for $I(X; Y)$ in terms of Shannon indices? We start by re-expressing the Shannon indices of the individual variables in terms of their probability distribu-

5 Techniques of Probability

tions:

$$I(X; Y) = -H(X, Y) - \sum_x p_X(x) \log p_X(x) - \sum_y p_Y(y) \log p_Y(y). \quad (5.95)$$

This, in turn, is the same as

$$\begin{aligned} I(X; Y) &= -H(X, Y) - \sum_{x,y} p(x, y) \log p_X(x) - \sum_{x,y} p(x, y) \log p_Y(y) \\ &= -H(X, Y) + \sum_{x,y} p(x, y) [-\log p_X(x) - \log p_Y(y)]. \end{aligned} \quad (5.96)$$

Using the definition of $H(X, Y)$ and the familiar properties of logarithms, we arrive at the *mutual information* between X and Y :

$$I(X; Y) = \sum_{x,y} p(x, y) \log \left(\frac{p(x, y)}{p_X(x)p_Y(y)} \right). \quad (5.97)$$

Note that $I(X; Y)$ vanishes if $p(x, y)$ factors neatly into $p_X(x)$ and $p_Y(y)$. In this case, knowing the value of x does not reduce the number of questions we require to ascertain the value of y , and vice versa.

Consider a set A of random variables, described by a joint probability distribution over all possible outcomes of all the variables in the set. The Shannon index can be shown to satisfy the following two properties.

- *Monotonicity*: The Shannon index of a subset $U \subset A$ that is contained in a subset $V \subset A$ cannot be larger than that of V . That is, if $U \subset V$, then $H(U) \leq H(V)$.
- *Strong subadditivity*: Given two subsets, the Shannon index of their union cannot exceed the Shannon index of each separately minus the index of their intersection:

$$H(U \cup V) \leq H(U) + H(V) - H(U \cap V). \quad (5.98)$$

Therefore, Shannon's formula is an information function which can be used in the multiscale structure formalism of Chapter 2.

Moreover, the Shannon index is a *natural* information function in the context of probabilities and random variables, because it is the unique functional which satisfies certain basic and fairly intuitive desiderata [225]. First, the value of information function applied to a probability distribution should not change if we expand the set of outcomes with new elements which are assigned probability zero. Second, an information function should be invariant under permutations of the probabilities: we are just as uncertain if our expectations are

$$p_0 = x, \quad p_1 = 1 - x, \quad (5.99)$$

as we are if

$$p_0 = 1 - x, \quad p_1 = x. \quad (5.100)$$

Third, for two experiments represented by random variables X and Y , an information function should always satisfy

$$H(X \cup Y) \leq H(X) + H(Y). \quad (5.101)$$

This follows from strong subadditivity, if we take $U = \{X\}$ and $V = \{Y\}$. Fourth, if the random variables X and Y are *independent*, we should have equality in the previous relation.

If we impose these four desiderata, then the only possible information functions are linear combinations of the Shannon index and the quantity

$$K[\{p_i\}] = \log_2 |\{p_i | p_i \neq 0\}|, \quad (5.102)$$

which counts the number of outcomes assigned greater than zero probability. This is known as the *Hartley entropy*. We can rule out a contribution of this form if we require in addition that for a random variable which has two possible outcomes, the information is small if the probability of one outcome is near zero.

Many other derivations of the Shannon index from sets of desiderata have been made. For a recent example, see Baez, Fritz and Leinster [226]. The characterization described here has the advantage that it relates fairly clearly to the axioms of the multiscale structure formalism.

5.5 Moments and Cumulants

The mean of a probability distribution $p(x)$ is

$$\langle x \rangle \equiv \sum_{x=-\infty}^{\infty} p(x)x,$$

or for a continuous distribution,

$$\langle x \rangle \equiv \int dx p(x)x.$$

This expression readily generalizes to that for the *expectation value* of a function f ,

$$\langle f(x) \rangle \equiv \int dx f(x)p(x). \quad (5.103)$$

Expectation values of powers of x are called *moments* of x :

$$\langle x^n \rangle = \int dx p(x)x^n. \quad (5.104)$$

Moments are interesting things to know about a probability distribution, but they are not always the most convenient values for characterizing how such distributions work.

5 Techniques of Probability

For example, suppose we have two random variables, X and Y , which independently of one another take values according to probability distributions $p_X(x)$ and $p_Y(y)$. What can we say about a third random variable $Z \equiv X + Y$? And, is there some characterization of p_X and p_Y such that we can just add properties of p_X and p_Y to obtain the corresponding property of p_Z ?

If Z takes the value z and X takes the value x , then Y must have taken the value $z - x$. Following the basic rules of probability, we deduce that because X and Y are independent, the probability of X taking the value x and Y the complementary value $z - x$ is $p_X(x)p_Y(z - x)$. But this is only one way for a measurement of Z to give the value z ; the sum $x + y$ can work out to z for any value of x . So, adding up the probabilities, we find

$$p_Z(z) = \int dx p_X(x)p_Y(z - x). \quad (5.105)$$

The probability distribution for the sum of two random variables is the *convolution* of the distributions for the random variables being added. This indicates that the higher moments of p_X and p_Y will not combine neatly when we construct p_Z .

A useful result, the *convolution theorem*, says that the Fourier transform of a convolution is the product of the Fourier transforms of the functions convolved, up to a constant depending on how the transformation was normalized. (The proof is both standard and fairly direct; see a good text on mathematical methods [227].) This suggests a way to proceed.

The *characteristic function* of a probability distribution is just its Fourier transform, which we can write as an expectation value:

$$\tilde{p}(k) \equiv \langle e^{-ikx} \rangle = \int dx p(x)e^{-ikx}. \quad (5.106)$$

Thus,

$$p(x) = \frac{1}{2\pi} \int dk \tilde{p}(k)e^{ikx}. \quad (5.107)$$

When we add random variables, their characteristic functions multiply, which means that the *logarithms* of their characteristic functions *add*. So, if we want to define quantities which are like moments but which add conveniently when we combine random variables, we should look at logarithms of characteristic functions.

Recall that

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots, \quad (5.108)$$

so we can expand the characteristic function as

$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n \right\rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle. \quad (5.109)$$

This relates the characteristic function to the moments.

As indicated, we're interested in logarithms of characteristic functions, because char-

5.5 Moments and Cumulants

characteristic functions multiply under convolution, and logarithms turn products into sums. Expanding the logarithm of $\tilde{p}(k)$ as we expanded $\tilde{p}(k)$ itself, we define the *cumulants* of x , written with the notation $\langle x^n \rangle_c$.

$$\log \tilde{p}(k) \equiv \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c. \quad (5.110)$$

When manipulating logarithms, it is sometimes useful to recall the Taylor expansion of the logarithm function for arguments near 1:

$$\log(1 + \epsilon) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\epsilon^n}{n}. \quad (5.111)$$

How do the cumulants relate to the moments?³ Can we find a simple formula for one set of numbers in terms of the other? We have two formulas involving the characteristic function $\tilde{p}(k)$, one using moments and the other using cumulants. If we equate the expressions for $\tilde{p}(k)$ in Eqs. (5.109) and (5.110), we can derive the relationship between the moments $\{\langle x^n \rangle\}$ and the cumulants $\{\langle x^n \rangle_c\}$. From Eqs. (5.109) and (5.110), we see that

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} \langle x^m \rangle &= \exp \left[\sum_{l=1}^{\infty} \frac{(-ik)^l}{l!} \langle x^l \rangle_c \right] \\ &= \prod_{l=1}^{\infty} \exp \left[\frac{(-ik)^l}{l!} \langle x^l \rangle_c \right] \\ &= \prod_{l=1}^{\infty} \sum_{n_l=0}^{\infty} \left[\frac{(-ik)^{ln_l}}{n_l!} \left(\frac{\langle x^l \rangle_c}{l!} \right)^{n_l} \right]. \end{aligned} \quad (5.112)$$

The powers of $(-ik)^m$ on both sides have to be equal, so

$$\frac{\langle x^m \rangle}{m!} = \sum_{\{n_l\}} \prod_l \frac{\langle x^l \rangle_c^{n_l}}{n_l! (l!)^{n_l}}, \quad (5.113)$$

in which the prime on the Σ means we are restricting the sum such that $\sum l n_l = m$. Moving the $m!$ to the right-hand side, we find

$$\langle x^m \rangle = \sum_{\{n_l\}} m! \prod_l \frac{1}{n_l! (l!)^{n_l}} \langle x^l \rangle_c^{n_l}. \quad (5.114)$$

The numerical factor multiplying $\langle x^l \rangle_c^{n_l}$ turns out to have a fairly simple combinatorial interpretation: it is the number of ways of breaking m points into clusters, satisfying

³This discussion is based on Kardar [49, 199], but with more intermediate steps worked out.

5 Techniques of Probability

the condition that for each value of l , the assignment has n_l clusters of l points.

This is easiest to see if we work out a simple example first. Suppose we have three points, which we wish to regroup into a set of two points and a third point left to itself. We can do this in three distinct ways, grouping together 1 and 2, 1 and 3 or 2 and 3. In other words, there exist three ways to assign a total of three particles to a 1-cluster and a 2-cluster.



If we write n_1 for the number of 1-clusters and n_2 for the number of 2-clusters in our decomposition, we can say that the number of ways W to organize our 3-point set is

$$W(n_1 = 1, n_2 = 1) = 3. \quad (5.115)$$

We now explore the behavior of W for a general set of m points. First, because every point has to be placed within one cluster or another, the cluster sizes times the number of clusters used have to add up to the total size of the set:

$$\sum_l l n_l = m. \quad (5.116)$$

We know that we can order an m -element set in $m!$ ways. We can find a general formula for W if we divide these $m!$ total permutations by the number of equivalent ones, that is, by the number of permutations which give indistinguishable results. Inside each subset of l elements, we can permute the labels in $l!$ different ways and still get an equivalent grouping. Furthermore, we can rearrange n_l subsets in $n_l!$ ways. The former gives us a factor of $(l!)^{n_l}$, and the latter a factor of $n_l!$. Thus,

$$W(\{n_l\}) = \frac{m!}{\prod_l n_l! (l!)^{n_l}}. \quad (5.117)$$

To sanity-check this formula, note that for $n_1 = n_2 = 1$,

$$W(1, 1) = \frac{3!}{(1!1!^1)(1!2!^1)} = \frac{3 \cdot 2}{2} = 3. \quad (5.118)$$

Putting everything together, we use this combinatorial notation to rewrite Eq. (5.114):

$$\langle x^m \rangle = \sum_{\{n_l\}} W(\{n_l\}) \prod_l \langle x^l \rangle_c^{n_l}, \text{ where } \sum_l l n_l = m. \quad (5.119)$$

Given any sequence $\{\langle x^l \rangle_c\}$ of cumulants, we can compute the moments. This means that if we have two independent random variables X and Y , described by the probability assignments p_X and p_Y , we can compute the moments of the new random variable

$X + Y$. We will apply this to very good effect in section §5.7.

It is a straightforward matter to work out the first few moments in terms of the first few cumulants. (And these relations are convenient to have on hand for reference.) First, because there's only one way to make a linked cluster out of one point, the first moment is the same as the first cumulant:

$$\langle x \rangle = \langle x \rangle_c. \quad (5.120)$$

The second moment, $\langle x^2 \rangle$, will have two terms, because we can make a diagram with two disconnected points (that is, two one-point clusters) and a diagram with two connected points (a 2-cluster).

$$\langle x^2 \rangle = \langle x^2 \rangle_c + \langle x \rangle_c^2 \quad (5.121)$$

The third moment $\langle x^3 \rangle$ is slightly more complicated, since we can chop up the three-vertex diagram into a 2-cluster and a 1-cluster in three ways. (We worked this out explicitly earlier.) This means that our formula will have a symmetry factor.

In diagram form,

$$\text{Three points} = \text{Three separate clusters} + 3 \times \text{Two-point cluster and one-point cluster} + \text{Three-point cluster} \quad (5.122)$$

An alternate way to draw this picture is as a forest of trees:

$$\text{Three points} = \text{Three separate rooted trees} + 3 \times \text{Two-child rooted tree and one-point rooted tree} + \text{Three-child rooted tree} \quad (5.123)$$

Here, the filled circles on the upper row indicate factors of x , and they belong to the same group if they are linked to the same circle on the bottom row. Each group, which stands for a cumulant, is a rooted tree, and each little forest is a term in the expansion of $\langle x^3 \rangle$. Either way, the equivalent in algebraic symbols is

$$\langle x^3 \rangle = \langle x^3 \rangle_c + 3 \langle x^2 \rangle_c \langle x \rangle_c + \langle x \rangle_c^3. \quad (5.124)$$

The expectation value of x^4 works analogously. Expanded out as products of cumulants, $\langle x^4 \rangle$ is a sum of five terms:

$$\langle x^4 \rangle = \langle x^4 \rangle_c + 4 \langle x^3 \rangle_c \langle x \rangle_c + 3 \langle x^2 \rangle_c^2 + 6 \langle x^2 \rangle_c \langle x \rangle_c^2 + \langle x \rangle_c^4. \quad (5.125)$$

Note that the total degree of each term on the right-hand side is equal to 4. If x is a quantity which carries units, this is necessary for dimensional consistency. The coefficients are combinatorial symmetry factors that come from the number of ways a set of four points can be partitioned. There are four ways to pull out a single point, three ways to divide the set into two pairs, and six ways to group together a pair of

5 Techniques of Probability

points while leaving the other two points out. As before, we have multiple equivalent ways to derive this: by counting ways to circle points, by counting our options to make forests out of trees, or by plugging into our algebraic formula for $W(\{n_i\})$. In fact, we can also build up these formulas recursively, by considering the operation of merging a new tree into a forest in all possible ways.

We have derived the first four moments in terms of the first four cumulants, using Eq. (5.119). Inverting these relationships yields the first four cumulants in terms of the first four moments:

$$\begin{aligned}\langle x \rangle_c &= \langle x \rangle \\ \langle x^2 \rangle_c &= \langle x^2 \rangle - \langle x \rangle^2 \\ \langle x^3 \rangle_c &= \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \\ \langle x^4 \rangle_c &= \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4.\end{aligned}\quad (5.126)$$

The first cumulant is just the mean. The second cumulant, the *variance*, is a measure of spread about the mean. Together, they are the cumulants referred to and made use of most regularly. Soon, in §5.7, we will see why this is the case. The relations in Eq. (5.126) are, in practice, about as high-order as one needs to calculate. Cumulants $\langle x^n \rangle_c$ with $n > 4$ aren't invoked as often.

We are now in a position to understand the approximations we made in §4.3.4. The only extra wrinkle is that in those calculations, we used *cross-cumulants* of multiple different random variables (for example, s_1 and s_2). This breaks the symmetry which yields the numerical coefficients, but the logic is otherwise the same:

$$\langle xyz \rangle = \langle xyz \rangle_c + \langle xy \rangle_c \langle z \rangle_c + \langle yz \rangle_c \langle x \rangle_c + \langle zx \rangle_c \langle y \rangle_c + \langle x \rangle_c \langle y \rangle_c \langle z \rangle_c. \quad (5.127)$$

We obtain the result we used in the previous chapter, Eq. (4.61), by imposing the approximation that the third-order cross-cumulant $\langle xyz \rangle_c$ vanishes. If the second-order cross-cumulants vanish as well, then the expectation value of the product is just the product of the means.

The coefficients in Eq. (5.125) show up somewhere else, too. Consider two functions, $f(x)$ and $g(x)$. Composing them yields a new function, $f(g(x))$, and the chain rule tells us its derivative:

$$\frac{d}{dx} [f(g(x))] = \frac{df}{dg} \frac{dg}{dx}. \quad (5.128)$$

Now, let's differentiate again. Using the product rule,

$$\frac{d^2}{dx^2} [f(g(x))] = \frac{d^2 f}{dg^2} \left(\frac{dg}{dx} \right)^2 + \frac{df}{dg} \frac{d^2 g}{dx^2}. \quad (5.129)$$

We can repeat this process, cranking through the algebra:

$$\frac{d^3}{dx^3} [f(g(x))] = 3 \left(\frac{dg}{dx} \right) \frac{d^2 f}{dg^2} \frac{d^2 g}{dx^2} + \left(\frac{dg}{dx} \right)^3 \frac{d^3 f}{dg^3} + \frac{df}{dg} \frac{d^3 g}{dx^3}. \quad (5.130)$$

And, taking it one more step,

$$\begin{aligned}
 \frac{d^3}{dx^3} [f(g(x))] &= \frac{d^4 f}{dg^4} \left(\frac{dg}{dx} \right)^4 \\
 &+ 6 \frac{d^3 f}{dg^3} \frac{d^2 g}{dx^2} \left(\frac{dg}{dx} \right)^2 \\
 &+ 3 \frac{d^2 f}{dg^2} \left(\frac{d^2 g}{dx^2} \right)^2 \\
 &+ 4 \frac{d^2 f}{dg^2} \frac{d^3 g}{dx^3} \frac{dg}{dx} \\
 &+ \frac{df}{dg} \frac{d^4 g}{dx^4}.
 \end{aligned} \tag{5.131}$$

This is sufficient to see a pattern take shape.

The coefficients in each sum are the same numbers, $W(\{n_l\})$, which we encountered when converting between moments and cumulants. In the ordinary chain rule, we have one term, and its coefficient is 1. For the second derivative, we get two terms, each of which has a coefficient of 1. In the third derivative, one term has a coefficient of 3, and when we go to the fourth derivative, the coefficients are the same as those in the expression for the moment $\langle x^4 \rangle$ in terms of the first four cumulants.

We can make the constructions exactly analogous if we identify a *connected cluster of k points* with the k^{th} derivative of $g(x)$. Each term is a derivative of f , times one or more factors which are derivatives of g . The diagram we used to demonstrate Eq. (5.115) also tells us the coefficients in the third derivative of $f(g(x))$.

Why should this be the case? We shall see in the next section.

5.6 Generating Functions

Our diagrammatic methods assign a *weight* to a graph, by interpreting that graph as an integral over a probability distribution. By construction, these graph weightings satisfy two properties. First, the weight of any graph is the product of the weights of its pieces, and second, the total weight of any collection of graphs is the sum of the weights of the individual graphs. These properties combine to yield an interesting and quite general result.

The moment $\langle x^n \rangle$ is the summed weight of all graphs having n vertices. To understand this series better, we “hang it on a clothesline” [228] by writing its *generating function*:

$$\mathcal{Q}(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \langle x^n \rangle. \tag{5.132}$$

Using Eq. (5.119), we can rewrite this generating function in terms of the cumulants

5 Techniques of Probability

$\langle x^n \rangle_c$:

$$\mathcal{Q}(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum'_{\{n_l\}} W(\{n_l\}) \prod_l \langle x^l \rangle_c^{n_l}, \text{ where } \sum_l n_l = n. \quad (5.133)$$

The l -th cumulant $\langle x^l \rangle_c$ is the weight of an l -vertex connected graph.

Because we are summing over all n , we can remove the restriction on the summation over $\{n_l\}$. First, we recall the definition of $W(\{n_l\})$.

$$\mathcal{Q}(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum'_{\{n_l\}} \frac{n!}{\prod_l n_l! (l!)^{n_l}} \prod_l \langle x^l \rangle_c^{n_l}. \quad (5.134)$$

One way to think of this is as follows: having “primed” the summation symbol means that we’ve inserted an implicit delta function. We can really sum over all sets $\{n_l\}$, as long as it’s understood that each term is multiplied by a Kronecker delta which is zero whenever $\sum_l n_l$ is not equal to n . But the outer summation symbol means that we have terms for all values of n , meaning that we can really take $\{n_l\}$ to be anything we like, as long as in that term, n is set equal to $\sum_l n_l$.

In frighteningly abstracted notation, we are saying that

$$\sum_{n=0}^{\infty} \sum_{\{n_l\}} \delta_{\sum_l n_l, n} = \sum_{\{n_l\}}, \quad (5.135)$$

so we can rewrite the generating function as

$$\begin{aligned} \mathcal{Q}(z) &= \sum_{\{n_l\}} z^{\sum_l n_l} \prod_l \frac{\langle x^l \rangle_c^{n_l}}{n_l! (l!)^{n_l}} \\ &= \sum_{\{n_l\}} \prod_l \frac{1}{n_l!} \left(\frac{z^l \langle x^l \rangle_c}{l!} \right)^{n_l}. \end{aligned} \quad (5.136)$$

Interchanging the product and summation symbols and employing the definition of the exponential,

$$\mathcal{Q}(z) = \prod_l \exp \left(z^l \frac{\langle x^l \rangle_c}{l!} \right), \quad (5.137)$$

or,

$$\boxed{\mathcal{Q}(z) = \exp \left(\sum_{l=1}^{\infty} \frac{\langle x^l \rangle_c}{l!} z^l \right)}. \quad (5.138)$$

The generating function over all graphs turns out to be the exponential of the generating function for *connected* graphs. This is known as the *linked-cluster theorem*.

We used only very general properties of graph weights in order to derive Eq. (5.138).

As long as those basic conditions are met, the same relationship will hold, no matter what interpretation we give to the vertices and edges of our graphs.

Another application occurs in statistical physics. If the vertices of our graphs represent particles and the graph weighting is determined by pairwise interactions between particles, then the generating function $\mathcal{Q}(z)$ has the form of the grand partition function, with the formal variable z playing the role of the fugacity $e^{\beta\mu}/\lambda^3$, which represents the difficulty of changing particle number by thermal fluctuations [49]. The grand partition function includes all “Mayer graphs”, connected or not; Eq. (5.138) states that the sum of all graphs is given by the exponential of the sum over connected graphs.

Now, we can see why the pattern of coefficients we saw in the iterated chain rule relates to moments and cumulants. If $f(z)$ is a generating function over some formal variable z ,

$$f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} f_n, \quad (5.139)$$

then we can extract the coefficient f_n by differentiating n times and evaluating the result at zero:

$$f_n = \left. \frac{d^n}{dz^n} f(z) \right|_{z=0}. \quad (5.140)$$

The differentiations remove all terms of lower order than n , and setting $z = 0$ removes all terms of higher order, leaving only f_n .

Consider the generating function defined by

$$f(z) = \exp[g(z)], \quad (5.141)$$

where $g(z)$ is itself a generating function,

$$g(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} g_n. \quad (5.142)$$

If we associate the expansion coefficients g_n with connected diagrams, then we are back to the linked-cluster theorem again. The values of f_n depend on the g_n . But we know that $f(z)$ is the *composition* of $g(z)$ with the exponential function, and we know that the f_n relate to the derivatives of $f(z)$, so we can compute the f_n by iterating the chain rule. The pictorial interpretation is that the operation of merging a new tree into a forest in all possible ways represents the product rule, which says that a derivative acts on a product to create a sum of new products, each of which has one factor modified. In probability theory, we also merge a new tree into a forest, when we go from the cumulant expansion of $\langle x^n \rangle$ to that of $\langle x^{n+1} \rangle$.

5.7 Central Limit Theorem

What happens when we add together many independent random variables? The same logic we used in §5.5 still applies: cumulants of their distributions add. If the random

5 Techniques of Probability

variable Y is given by

$$Y = \sum_{i=1}^N X_i, \quad (5.143)$$

where X_i are independent random variables described by distributions $p(x_i)$, then the n th cumulant of Y , $\langle y^n \rangle_c$, is simply the sum

$$\langle y^n \rangle_c = \sum_{i=1}^N \langle x_i^n \rangle_c. \quad (5.144)$$

To take the simplest case, suppose that the distributions of all the X_i are identical. Then $\langle y^n \rangle_c = N \langle x^n \rangle_c$, meaning that the new random variable

$$Z \equiv \frac{Y - N \langle x \rangle_c}{\sqrt{N}} \quad (5.145)$$

has a mean of zero, and higher cumulants which scale as $\langle z^n \rangle_c \propto N^{1-n/2}$. For N sufficiently big, all cumulants higher than the second die off, and Z becomes *Gaussian*.

$$\boxed{\lim_{N \rightarrow \infty} p \left(z = \frac{\sum_i x_i - N \langle x \rangle_c}{\sqrt{N}} \right) = \frac{1}{\sqrt{2\pi \langle x^2 \rangle_c}} \exp \left(-\frac{z^2}{2 \langle x^2 \rangle_c} \right)}. \quad (5.146)$$

We have arrived at the *Central Limit Theorem*: the sum of many uncorrelated random variables is itself a random variable, whose probability distribution is approximately Gaussian. A Gaussian distribution is specified by two parameters, which we denote here as μ and σ :

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right). \quad (5.147)$$

The Fourier transform of a Gaussian curve is itself a Gaussian:

$$\tilde{p}(k) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{-ikx} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) = \exp \left(-ik\mu - \frac{k^2\sigma^2}{2} \right). \quad (5.148)$$

We can read off the cumulants:

$$\langle x \rangle_c = \mu, \quad \langle x^2 \rangle_c = \sigma^2, \quad (5.149)$$

and $\langle x^n \rangle_c = 0$ for all $n > 2$. The Gaussian distribution is characterized entirely by its first two cumulants, namely the mean μ and the variance σ^2 .

A neat thing happens when we take Eqs. (5.120) through (5.125) and set all cumu-

5.7 Central Limit Theorem

lants of higher order than the variance to zero. Carrying this out, we get,

$$\begin{aligned}\langle x \rangle &= \mu \\ \langle x^2 \rangle &= \sigma^2 + \mu^2 \\ \langle x^3 \rangle &= 3\sigma^2\mu + \mu^3 \\ \langle x^4 \rangle &= 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4.\end{aligned}\tag{5.150}$$

In diagram language, our graphs are allowed to have clusters of one point and clusters of two points, but clusters of three or more points cause the graph value to vanish. Alternatively, in the forest picture, trees with more than two leaf nodes evaluate to zero. The numerical value of a forest is the product of the values of its constituent trees, so any forest containing a tree with too many leaf nodes will, likewise, contribute nothing to the sum total.

Recall the approximation we made in §4.3.4, in order to understand random neutral drift (and, thus, weak-selection evolutionary dynamics) on the hexagonal lattice. As explained above, that approximation involves setting cross-cumulants of third and higher order to zero. We can, therefore, consider it a *Gaussian closure*.

One of the many applications of the Central Limit Theorem is to *random walks*. Consider a walker that starts at $x = 0$ takes steps whose direction and length are chosen at random, according to a probability distribution $p_1(l)$. We can choose, conventionally, that negative increments are motion toward the left, and positive increments are steps toward the right. What is the probability to be at position x after N steps? How do we characterize this probability distribution, and how does it change over time, as the walker takes more steps?

The position after N steps is

$$x = l_1 + l_2 + \cdots + l_N.\tag{5.151}$$

Because the l_i come from identical independent distributions, the average displacement is

$$\langle x \rangle = \langle l_1 \rangle + \cdots + \langle l_N \rangle = N \langle l \rangle.\tag{5.152}$$

However, the final displacement of the walker can be greater or smaller than this value. The probability distribution for the net displacement will have some variance,

$$\langle x^2 \rangle_c = \sum_{i,j=1}^N (\langle l_i l_j \rangle - \langle l_i \rangle \langle l_j \rangle)\tag{5.153}$$

$$= \sum_{i,j=1}^n \langle l^2 \rangle_c \delta_{ij}\tag{5.154}$$

$$= N \langle l^2 \rangle_c.\tag{5.155}$$

All the cumulants follow this general pattern: after N steps, their value is multiplied by N . This is a specific example of the general property which we established for

5 Techniques of Probability

cumulants of independent random variables added together. By the Central Limit Theorem, we can neglect the higher-order cumulants for $N \gg 1$. Thus, the final probability distribution for large N will be a Gaussian, which we can write

$$p(x, N) = \exp \left[-\frac{(x - N \langle l \rangle)^2}{2N \langle l^2 \rangle_c} \right] \frac{1}{\sqrt{2\pi N \langle l^2 \rangle_c}}. \quad (5.156)$$

Because $p(x, N)$ is an exponential function, taking its derivative yields back $p(x, N)$ again. We work out the first two derivatives with respect to position, and the first derivative with respect to N :

$$\frac{\partial p}{\partial x} = -\frac{x - N \langle l \rangle}{N \langle l^2 \rangle_c} p, \quad (5.157)$$

$$\frac{\partial^2 p}{\partial x^2} = \frac{(x - N \langle l \rangle)^2}{N^2 \langle l^2 \rangle_c^2} p - \frac{1}{N \langle l^2 \rangle_c} p, \quad (5.158)$$

and, formally,

$$\frac{\partial p}{\partial N} = \left[\frac{\langle l \rangle (x - N \langle l \rangle)}{N \langle l^2 \rangle_c} + \frac{(x - N \langle l \rangle)^2}{2N^2 \langle l^2 \rangle_c} - \frac{1}{2N} \right] p. \quad (5.159)$$

We can use the first two formulae to simplify the third:

$$\frac{\partial p}{\partial N} = -\langle l \rangle \frac{\partial p}{\partial x} + \frac{\langle l^2 \rangle_c}{2} \frac{\partial^2 p}{\partial x^2}. \quad (5.160)$$

Instead of expressing how p changes with the number of steps N , we can say how p changes over time, if we introduce a timestep τ . Define

$$v = \frac{\langle l \rangle}{\tau}, \quad D = \frac{\langle l^2 \rangle_c}{\tau}. \quad (5.161)$$

Then, we obtain a *diffusion equation*, which relates the spatial and temporal partial derivatives of the probability density:

$$\frac{\partial p(x, t)}{\partial t} = -v \frac{\partial p}{\partial x} + \frac{D}{2} \frac{\partial^2 p}{\partial x^2}, \quad (5.162)$$

If the random walk is *unbiased*, then $v = 0$, and the diffusion equation simplifies to

$$\boxed{\frac{\partial p(x, t)}{\partial t} = \frac{D}{2} \frac{\partial^2 p}{\partial x^2}}. \quad (5.163)$$

We derived this diffusion equation in terms of the probability that a single random walker will arrive at position x . The equation codifies how our expectations for the walker at different times relate to each other. If we have many walkers moving along the same axis but not interacting with each other, then the fraction of those walkers which we expect to fall within a given interval scales with the integral of the proba-

bility density $p(x, t)$ over that interval. That is, the probability density governs our expectation for the number density.

This picture would become significantly more involved if the different random walkers in the population interacted with one another. In that case, we would have no reason to assume that the joint probability distribution over all walker positions should factor nicely into a product of single-walker probability densities.

The left-hand side of the diffusion equation (5.163) is a time derivative, so its units are those of p divided by time. On the right-hand side, we have a second derivative with respect to position, which introduces units of length^{-2} . In order for the dimensions on both sides to match, the constant D must have units of length^2 per time. This is one way to remember the intuition that for a diffusive process, variance grows linearly with time, or to say it another way, a characteristic length scale increases with the square root of the time elapsed.

5.8 Variations on Diffusion

Diffusion becomes more complicated when multiple types of particle are moving together through the same space, and the overall amount of space available to move through is limited. By considering the effect of volume limitations at the rate-equation level, we can derive in a simpler way some interesting modifications to the ordinary diffusion equation, alterations which have been seen to emerge from a highly involved analysis [132, 133].

We consider three sites, with occupation numbers a_1 , a_2 and a_3 . Particles jump away from a site at an overall rate k , and jumps are equally likely in either direction. We make the approximation that the *actual* flow is the same as the *most probable* flow, so we can quantify everything in terms of average rates. The change of the population at site 2 is a result of flows into site 2 from sites 1 and 3, and the flows out of site 2 into its neighbors:

$$\frac{da_2}{dt} = \frac{k}{2}a_1 - ka_2 + \frac{k}{2}a_3. \quad (5.164)$$

We rearrange this as follows:

$$\frac{da_2}{dt} = \frac{k}{2}(a_1 - 2a_2 + a_3) \quad (5.165)$$

$$= \frac{k}{2}(a_1 - a_2 + a_3 - a_2) \quad (5.166)$$

$$= \frac{k}{2}[(a_3 - a_2) - (a_2 - a_1)]. \quad (5.167)$$

The time derivative of a_2 is the difference of two differences, which we can express as

$$\frac{da_2}{dt} = \frac{k}{2}[(\Delta a)_2 - (\Delta a)_1], \quad (5.168)$$

5 Techniques of Probability

or in terms of the *discrete Laplacian operator* as

$$\boxed{\frac{da_2}{dt} = \frac{k}{2}(\Delta^2 a)_2.} \quad (5.169)$$

This is exactly the discrete analogue of the diffusion equation for a continuous line, Eq. (5.163).

A more intricate case is the diffusion of two substances through a common space, with a limit imposed on how much total substance can be at any point:

$$a_i + b_i \leq N, \quad \forall i. \quad (5.170)$$

Writing e_i for the amount of empty space at location i , we can turn this inequality into an equality:

$$a_i + b_i + e_i = N. \quad (5.171)$$

Because flow can only happen into empty space, the population of type a at site 2 changes as

$$\frac{da_2}{dt} = -k(e_1 + e_3)a_2 + \frac{k}{2}a_1e_2 + \frac{k}{2}a_3e_2. \quad (5.172)$$

Solving for e_i in terms of a_i and b_i , this becomes

$$\frac{da_2}{dt} = \frac{kN}{2}(a_1 - 2a_2 + a_3) + \frac{k}{2}[(b_1 + b_3)a_2 - b_2(a_1 + a_3)]. \quad (5.173)$$

The first term, which carries a factor of N , has the same functional dependence on the $\{a_i\}$ that we saw before, with the ordinary diffusion process. The second term is more complicated, having an interdependence between the $\{a_i\}$ and the $\{b_i\}$. A bit of algebra reveals a convenient way to state that interdependence:

$$(b_1 + b_3)a_2 - b_2(a_1 + a_3) = a_2(b_1 + b_3) - 2a_2b_2 - b_2(a_1 + a_3) + 2b_2a_2 \quad (5.174)$$

$$= a_2(b_1 - 2b_2 + b_3) - b_2(a_1 - 2a_2 + a_3). \quad (5.175)$$

This has the structure of “self times the Laplacian of the other, minus the other times the Laplacian of the self.” The same holds true for the time evolution of b_i , with the roles of the a - and b -variables interchanged.

Now, we express the $\{a_i\}$ and the $\{b_i\}$ as *fractions* of the volume at each site, N :

$$a_i = Nf_i, \quad b_i = Ng_i. \quad (5.176)$$

The time derivative of a_2 tells us the time derivative of f_2 , which in terms of the f - and g -variables works out to be

$$\frac{df_2}{dt} = \frac{kN}{2}(f_1 - 2f_2 + f_3) + \frac{kN}{2}[(g_1 + g_3)f_2 - g_2(f_1 + f_3)]. \quad (5.177)$$

Using the discrete Laplacian operator,

$$\boxed{\frac{df_i}{dt} = \frac{kN}{2}(\Delta^2 f)_i + \frac{kN}{2} [f_i(\Delta^2 g)_i - g_i(\Delta^2 f)_i]}. \quad (5.178)$$

The differential equation we had before is modified by *cross-diffusion* terms, which couple the time evolutions of the two fields. These new terms arise from the fact that limiting the total available volume at each site curtails the maximum possible flow rate.

What if spreading happens by *reproduction* instead of by hopping? This is, for example, characteristic of the spatial host–consumer model we studied at length in Chapter 3, and the evolutionary game-based lattice model of Chapter 4. As before, we require empty space to move into, so the growth rate from site i into site j should go like

$$ka_i e_j = ka_i(N - a_j - b_j). \quad (5.179)$$

There is no hopping out of a site, only the possibility of budding into it. So, the rate of change at site a_2 must be given by the rate at which budding events can happen, which depends upon the number of particles present at the neighboring sites, modulated by the amount of empty space available at a_2 . Implementing this idea in equations,

$$\frac{da_2}{dt} = ka_1(N - a_2 - b_2) + ka_3(N - a_2 - b_2) \quad (5.180)$$

$$= k(N - a_2 - b_2)(a_1 + a_3) \quad (5.181)$$

$$= k(N - a_2 - b_2) [(\Delta^2 a)_2 + 2a_2]. \quad (5.182)$$

Examining this result, we see an *effective diffusion* term appear: as before, the time derivative of a_2 can be written using $(\Delta^2 a)_2$.

Cross-diffusion terms like those in Eq. (5.178) become important in the study of *reaction-diffusion* systems. Suppose we have two substances, whose densities are given by the functions $a(x, t)$ and $b(x, t)$. The particles of both substances can spread diffusively, but when they bump into each other, reactions can take place, which introduce the possibility of creating or destroying particles. How many such reactions happen depends on the concentration of particles present. So, we write a system of equations to express how the densities should change:

$$\frac{\partial a(x, t)}{\partial t} = D_1 \frac{\partial^2 a}{\partial x^2} + f_1(a, b), \quad (5.183)$$

$$\frac{\partial b(x, t)}{\partial t} = D_2 \frac{\partial^2 b}{\partial x^2} + f_2(a, b). \quad (5.184)$$

It is a common practice to start with a model defined with a single location, writing functions f_1 and f_2 to represent the local dynamics, and then promote this construction to a spatial model by making a and b position-dependent and adding the diffusion terms. However, if the effective available volume at each position is limited, then this

5 Techniques of Probability

is not correct [132, 133]. We have in that case to include the cross-diffusion terms:

$$\frac{\partial a(x, t)}{\partial t} = D_1 \left[\frac{\partial^2 a}{\partial x^2} + a \frac{\partial^2 b}{\partial x^2} - b \frac{\partial^2 a}{\partial x^2} \right] + f_1(a, b), \quad (5.185)$$

$$\frac{\partial b(x, t)}{\partial t} = D_2 \left[\frac{\partial^2 b}{\partial x^2} + b \frac{\partial^2 a}{\partial x^2} - a \frac{\partial^2 b}{\partial x^2} \right] + f_2(a, b). \quad (5.186)$$

The cross-diffusion terms introduce an extra interdependence between the $a(x, t)$ and $b(x, t)$, above and beyond that which might be defined in the reaction terms $f_1(a, b)$ and $f_2(a, b)$.

6 Stochastic Adaptive Dynamics

In Chapter 4, we examined the dynamics of evolutionary games with a discrete set of strategies, and we wrote coupled differential equations to define dynamical systems. The competing populations in our dynamical systems were continuously variable in size. Now, we consider a complementary type of scenario, in which the game-players' strategies themselves are continuously variable. For example, the amount to which an individual is willing to contribute to a group-level effort to attain some social good could be a continuous quantity.

To simplify our analysis, we shall follow a precedent set in the literature and focus on cases where the total population size is constant. This is another way in which the current chapter is complementary to Chapter 4. The models we study in this chapter belong to an area of evolutionary theory known as *adaptive dynamics*. Our goal will be to extend recent results in adaptive dynamics beyond the deterministic limit and into a stochastic regime.

6.1 Justifying the Fokker–Planck Equation

Consider a random walk over a set of sites, in which the walker moves from site i to site j at a rate Π_{ji} . That is, the probability to jump in a short time dt is $\Pi_{ji}dt$. How do the occupation probabilities $\{P_i(t)\}$ change over time? A given occupation probability can change because the walker is likely to jump *to* that site or because a walker at that site is likely to jump *away*. In mathematics,

$$\frac{dP_i}{dt} = - \sum_j \Pi_{ji} P_i + \sum_j \Pi_{ij} P_j. \quad (6.1)$$

This relation is known as the *master equation*. By taking the continuum limit, in which the sites are very closely squished, we can arrive at another useful relation. We essentially make the transformations

$$i \rightarrow x, \quad P_i \rightarrow P(x, t), \quad \Pi_{ji} \rightarrow \Pi(x' - x, x). \quad (6.2)$$

Note that we have re-parametrized the jump probability to use the difference of positions, rather than the positions themselves. (The two parametrizations are, of course, entirely equivalent.) The result of transforming Eq. (6.1) in this fashion is

$$\frac{\partial}{\partial t} P(x, t) = - \int dx' \Pi(x' - x, x) P(x, t) + \int dx' \Pi(x - x', x') P(x', t). \quad (6.3)$$

6 Stochastic Adaptive Dynamics

The transition rates depend on the separation between the old position and the new position, $y = x' - x$. A small change in the new position x' implies a small change in the separation: $dx' = dy$. Using the new variable y , Eq. (6.3) becomes

$$\frac{\partial}{\partial t} P(x, t) = - \int dy \Pi(y, x) P(x, t) + \int dy \Pi(y, x - y) P(x - y, t). \quad (6.4)$$

For most physical applications, we expect typical changes to be local, *i.e.*, that Π is dominated by small y . We can then expand the second integral in Eq. (6.4) as a Taylor series. The first term in this expansion cancels with the first integral in Eq. (6.4). The next two terms yield

$$\frac{\partial}{\partial t} P(x, t) = - \int dy y \frac{\partial}{\partial x} (\Pi(y, x) P(x, t)) + \frac{1}{2} \int dy y^2 \frac{\partial^2}{\partial x^2} (\Pi(y, x) P(x, t)). \quad (6.5)$$

We can take the derivatives outside of the integrals, like so:

$$\frac{\partial}{\partial t} P(x, t) = - \frac{\partial}{\partial x} \left[P(x) \int dy y \Pi(y, x) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[\int dy y^2 (\Pi(y, x) P(x, t)) \right]. \quad (6.6)$$

This becomes a close analogue of the diffusion equation, which we derived in Eq. (5.163). Each term involves a factor which is essentially a moment of $\Pi(y, x)$ with respect to y . Define

$$F(x) = \int dy y \Pi(y, x), \quad (6.7)$$

and

$$D(x) = \int dy y^2 \Pi(y, x). \quad (6.8)$$

Then

$$\boxed{\frac{\partial P(x, t)}{\partial t} = - \frac{\partial}{\partial x} [F(x) P(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [D(x) P(x, t)]}. \quad (6.9)$$

Eq. (6.9) is known as the *Fokker-Planck equation*. For many applications, $D(x)$ can be taken as a constant, and so,

$$\frac{\partial P(x, t)}{\partial t} = - \frac{\partial}{\partial x} [F(x) P(x, t)] + \frac{D}{2} \frac{\partial^2 P(x, t)}{\partial x^2}. \quad (6.10)$$

An appropriate change of variables can in fact transform away the position dependence of $D(x)$, so this special case can be used even when the ease of diffusion is position-dependent [229].

If we apply the product rule, we can expand the derivatives in Eq. (6.9), obtaining

$$\frac{\partial P(x, t)}{\partial t} = - \frac{D}{2} P'' + (D' - F) P' + \left(\frac{D''}{2} - F' \right) P. \quad (6.11)$$

Note that this expression includes both the first derivative of $F(x)$ and the second

6.1 Justifying the Fokker–Planck Equation

derivative of $D(x)$. This will be important soon.

To find the steady-state solution, $P^*(x)$, for the Fokker–Planck equation with constant diffusion, set the derivative with respect to time equal to zero. This implies

$$\frac{\partial}{\partial x} [F(x)P^*(x)] = \frac{D}{2} \frac{\partial^2 P^*(x)}{\partial x^2}, \quad (6.12)$$

or

$$\frac{\partial^2 P^*(x)}{\partial x^2} = \frac{2}{D} \frac{\partial}{\partial x} [F(x)P^*(x)]. \quad (6.13)$$

Integrating once over x ,

$$\frac{\partial P^*(x)}{\partial x} = \frac{2F(x)}{D} P^*(x). \quad (6.14)$$

When the derivative of a function gives back the function itself, the function is an exponential. The question is, an exponential of what? The derivative of its argument with respect to x must be $2F(x)/D$. Recall that the derivative of an integral with respect to its upper limit is the integrand evaluated at that limit. So,

$$P^*(x) \propto \exp \left(\frac{2}{D} \int^x dx' F(x') \right). \quad (6.15)$$

The choice of the lower limit and the constant of integration glossed over earlier can be absorbed into the prefactor which establishes the proper normalization,

$$\int dx P^*(x) = 1. \quad (6.16)$$

If the diffusion function $D(x)$ is not constant, then we can find the steady-state distribution in much the same way. The result is the slightly modified formula

$$P^*(x) \propto \frac{1}{D(x)} \exp \left(2 \int^x dx' \frac{F(x')}{D(x')} \right). \quad (6.17)$$

In what follows, we will use the Fokker–Planck equation in the following way. The variable x will denote the *value of a genetic trait in a population*. We will assume that the population is uniform: each time we survey it, all the organisms have the same genotype. However, in between observations, mutations can occur, creating individuals with slightly different genotypes. If the offspring of a mutant take over the population, then the value of x will change by a small amount. If the mutants fail to supplant the native population, then x stays the same. The timescale of this ecological competition will, in our analysis, be much shorter than the timescale over which we follow the variable x .

Even though the population is genetically uniform at each observation, we might not be *certain* about its genetic composition. For example, in a practical context we might only be able to carry out a few observations, and the measurements we take might be confounded by environmental factors, meaning that multiple values of x could be consistent with the data we gather. We therefore summarize our knowledge

of the population by a probability distribution over x . If we know how mutations can happen, then we can say something about what x might be at a later time. However, because the outcomes of mutation events and competition are not certain, our statements about what x might be in the future must also be probabilistic. The function $P(x, t)$ expresses what we can deduce about what x might be at time t , given the information currently at hand. When the time t actually rolls around, we might go out and gather new data, allowing us to refine our information about what genotypes could be present. (This is the kind of problem for which we derived probability-update rules in Chapter 5.) For the most part, we will not be concerned here with that side of the problem; we will in this chapter focus on the computation of expectations in the absence of new information.

6.2 The Deterministic Limit

We define an evolutionary process in the following way: at any time, a mutant can arise in an otherwise uniform population. The mutation rate can in general depend upon the current trait value, and the mutated trait value is near that of the resident population, but displaced by a random small amount. Whether the mutant variety takes over the population or not depends upon how the two varieties interact, which we codify as a game. The payoff for a mutant individual with trait value x_M playing against another individual drawn from a population having trait value x_R is $A(x_M; x_R)$.

An article by Allen, Nowak and Dieckmann [230] studies this scenario in considerable detail, and for the remainder of this section, we follow their notation fairly closely.

As time progresses, the population's trait value can change. Let

$$a_{ij} = A(x_i; x_j), \quad i, j \in \{M, R\}. \quad (6.18)$$

We write the payoff matrix for mutant- and resident-type individuals interacting via two-player games as

$$G = \begin{pmatrix} a_{MM} & a_{MR} \\ a_{RM} & a_{RR} \end{pmatrix}. \quad (6.19)$$

The probability that a variety with trait value x' takes over a population with trait value x is $\rho(x'; x)$, which is some function of the matrix G .

Because we are considering two-player games with a set of two strategies, we can make use of a considerable amount of theory that has been developed for that case. It can be proved that, in a two-player game with two strategies, if the effect of selection is not too strong, then a strategy R is favored over another strategy M if the following inequality is satisfied:

$$\sigma(a_{RR} - a_{MM}) + a_{RM} - a_{MR} > 0. \quad (6.20)$$

Here, σ denotes the Tarnita *structure coefficient*, a number that depends on the population structure and the update rule, but is independent of the payoff matrix [231]. If this inequality is satisfied, then the fixation probability of R is greater than that

6.2 The Deterministic Limit

of M : a single R -type individual in a population of type M is more likely to take over the ecosystem than a single M -type individual in the reverse scenario. Structure coefficients have been calculated for several different cases of interest [197, 230, 231].

Denote the mutation rate per individual by $u(x)$, and let the mutation step size be described by a probability density $\mu(z)$, which we take to be centered at zero. We assume that $\mu(z)$ is a narrow distribution, such that the expectation value of $|z|^2$ is much greater than that of $|z|^3$. Because many of our expressions will involve the variance of $\mu(z)$, we'll abbreviate it as ν :

$$\nu = \int dz z^2 \mu(z). \quad (6.21)$$

Suppose the current trait value is established to be x . By how much can we expect the trait value to change? We find this by integrating over the step size z :

$$\langle \Delta x \rangle = \int dz z N u(x) \mu(z) \rho(x+z; x) = N u(x) \int dz z \mu(z) \rho(x+z; x). \quad (6.22)$$

If we expand the probability $\rho(x+z; x)$ to first order in the jump distance z ,

$$\rho(x+z; x) = \rho(x; x) + z \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x} + \mathcal{O}(|z|^2), \quad (6.23)$$

then our expectation value $\langle \Delta x \rangle$ becomes

$$\begin{aligned} \langle \Delta x \rangle &= N u(x) \rho(x; x) \int dz z \mu(z) \\ &\quad + N u(x) \int dz z^2 \mu(z) \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x} \\ &\quad + \mathcal{O} \left[\int dz \mu(z) |z|^3 \right]. \end{aligned} \quad (6.24)$$

The first term vanishes by symmetry, and the third is negligible by assumption. Therefore,

$$\langle \Delta x \rangle = N u(x) \int dz z^2 \mu(z) \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x} = N u(x) \nu \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x}. \quad (6.25)$$

This relates the expected change in x to the derivative of the fixation probability $\rho(x'; x)$.

The simplest way to turn this into a dynamical system is to impose the condition that the *actual* change in x is given by this *expected* change in x . Doing so, we arrive at the *canonical equation* for the time evolution of the trait value:

$$\boxed{\frac{dx}{dt} = N u(x) \nu \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x}}. \quad (6.26)$$

The derivative on the right-hand side can be written using the chain rule as

$$\left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x} = \frac{1}{A(x; x)} \sum_{j,k} \left. \frac{\partial \rho}{\partial a_{jk}} \right|_{G=J} \left. \frac{\partial a_{jk}}{\partial x'} \right|_{x'=x}. \quad (6.27)$$

6.3 A Fokker–Planck Equation for Adaptive Dynamics

Having derived Eq. (6.26) for the deterministic limit, we now push our understanding into new territory by including the effects of stochastic fluctuations. Instead of a single coordinate value which depends on time, $x(t)$, we will consider a time-varying probability density over the possible population states, $P(x, t)$.

In the previous section, we equated expected change in x with actual change, thereby establishing a deterministic dynamic. Now, we relax that requirement. Our calculation of the expected change, Eq. (6.25), is still valid, but it no longer tells the full story. Looking back over our derivation of the Fokker–Planck equation, we see that need a quantity which represents how much the probability density $P(x, t)$ spreads out over time. If we define

$$\langle (\Delta x)^2 \rangle = \int dz z^2 u(x) \mu(z) \rho(x + z; x), \quad (6.28)$$

then we find that

$$\langle (\Delta x)^2 \rangle = u(x) \nu \rho(x; x). \quad (6.29)$$

We can use Eqs. (6.25) and (6.29) to construct a Fokker–Planck equation for the time evolution of $P(x, t)$. The result is

$$\boxed{\frac{\partial P(x, t)}{\partial t} = - \frac{\partial}{\partial x} \left[u(x) \nu \left. \frac{\partial \rho(x'; x)}{\partial x'} \right|_{x'=x} P(x, t) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [u(x) \nu \rho(x; x) P(x, t)]}. \quad (6.30)$$

Note that the time evolution of the probability density depends on *second* derivatives of ρ . In contrast, the deterministic system defined by the canonical equation, Eq. (6.26), depends only on *first* derivatives of ρ .

We said that the game which governs the invasion dynamics is a two-player interaction. What if it is a multiplayer game instead? Recently, Allen, Nowak and Dieckmann [230] proved that for the deterministic system based on Eq. (6.26), a multiplayer interaction is actually equivalent to a two-player game. If the payoff function for the multiplayer game is $B(x; y_1, y_2, \dots, y_{n-1})$, and the game is symmetric under permutations of the y -arguments, then the dynamics are equivalent to those of a model defined in terms of the pairwise game

$$A(x; y) = B(x; y, y, \dots, y). \quad (6.31)$$

However, their proof depends crucially on the fact that the time-evolution equation uses only first derivatives of ρ . This means that it does not apply in the general stochastic context.

6.3 A Fokker–Planck Equation for Adaptive Dynamics

The Fokker–Planck equation (6.30), which we can think of as a “canonical diffusion equation for adaptive dynamics,” is equivalent to the stochastic differential equation derived by Champagnat and Lambert [232]. The route we have taken is more in line with a physicist’s approach to stochastic processes. A similar derivation was recently given by Van Cleve [233].

From the steady-state solution to the Fokker–Planck equation, (6.17), we can derive the mutation-selection equilibrium for stochastic adaptive dynamics:

$$P^*(x) \propto \frac{1}{u(x)\nu\rho(x;x)} \exp\left(2 \int^x dx' \frac{1}{\rho(x';x')} \frac{\partial \rho(y;x')}{\partial y} \Big|_{y=x'}\right). \quad (6.32)$$

The integrand in the exponential includes a first derivative of ρ , but not any higher derivatives. We might be able to rescue the theorem of Allen, Nowak and Dieckmann! Specifically, we might be able to relate the mutation-selection equilibrium distribution for a multiplayer game to that of a pairwise game, even if the *approach* to equilibrium does not match. However, before we can do that, we have to understand $\rho(x;x)$, the probability that a mutant organism with the same genotype as the resident population can take over the ecosystem by genetic drift. This depends on the population structure and the update rules. So, in order to reduce multilateral interactions to pairwise ones, we have to construct an appropriate structure of pairwise relationships and rules of succession.

That is a very general problem. For the moment, we focus instead on a more specific one, zeroing our attention in on a particular two-player game. In Chapters 2 and 4, we examined the Prisoner’s Dilemma, implementing the basic idea of it in two different settings. Can we do the same here? Within the world of adaptive dynamics, strategies are continuously variable, rather than binary. And indeed, simplifying multiplayer games to dyadic ones hinges upon that smooth variation.

We can write the payoff function for a continuous Prisoner’s Dilemma as

$$A(x;y) = -C(x) + B(y). \quad (6.33)$$

The value $C(x)$ is the cost which an individual having genotype x pays in order to benefit another, and the amount of benefit which the other obtains is given by the function B . We set

$$C(0) = B(0) = 0, \quad (6.34)$$

and we require that both $C(x)$ and $B(x)$ are strictly increasing, with the inequality $C(x) < B(x)$ satisfied for $x > 0$. For the mathematics to work smoothly, we also posit that $C(x)$ and $B(x)$ are twice differentiable.¹

Assuming that the neutral-drift takeover probability can be given in terms of an

¹We note here a difference between our approach and that of Van Cleve [233]. In this chapter, the continuously variable trait x is a *strength of cooperation*, while in Van Cleve’s study, the evolvable trait is the *fraction of time* in which an agent cooperates in a *discrete* game.

6 Stochastic Adaptive Dynamics

effective population size N_e , we find that

$$P^*(x) \propto \frac{N_e}{u(x)\nu} \exp \left(2N_e \int_0^x dx' \frac{\partial \rho(y; x')}{\partial y} \Big|_{y=x'} \right). \quad (6.35)$$

Plugging in the Prisoner's Dilemma, this becomes

$$P^*(x) \propto \frac{N_e}{u(x)\nu} \exp \left(2N_e \int_0^x \frac{dx'}{A(x'; x')} \left[-C'(x') + \frac{\sigma - 1}{\sigma + 1} B'(x') \right] \right), \quad (6.36)$$

where σ is the Tarnita structure coefficient, which as we said earlier depends on the population structure and on the update rule [231]. To simplify matters, take the mutation rate $u(x)$ to be constant. Carrying out the integral in the exponential is difficult; however, we can understand a great deal about the solution by using the fact that the exponential will be peaked where its argument is maximized. To find this extremum, we can differentiate the argument with respect to x , and by the Fundamental Theorem of Calculus, this just extracts the integrand, evaluated at $x' = x$.

If $\sigma \leq 1$, then the argument of the exponential is always nonpositive, so the probability density $P^*(x)$ piles up at $x = 0$. On the other hand, if $\sigma > 1$, then $P^*(x)$ is peaked at the value of x which satisfies

$$\boxed{\frac{B'(x)}{C'(x)} = \frac{\sigma + 1}{\sigma - 1}}. \quad (6.37)$$

Ratios of benefits to costs are commonplace in the study of how social behaviors evolve. Recall that in Chapter 4, we found the stability criterion for the nonspatial Volunteer's Dilemma in terms of just such a ratio. We will see similar ratios again in Chapters 8 and 9. Eq. (6.37) is an interesting variant, in that it is a comparison of *marginal* quantities, *i.e.*, of the derivatives of the benefit and cost functions. It tells us that the peak of $P^*(x)$ depends on the population structure and the details of the organisms' life cycles, through the structure coefficient σ .

Note that if the mutation rate u is not constant but instead varies with x , then the shape of the mutation-selection equilibrium curve $P^*(x)$ will change. This is another example of a theme noted by Allen and Tarnita [33]: standards of evolutionary success depend upon mutation rates. We saw this issue arise in Chapter 2, where the mutation rate had a strong effect upon the shape of the mutation-selection equilibrium curve, and in turn, upon the complexity profile. It is a general, but underappreciated, phenomenon [234].

Another scenario of evolutionary interest is the *Snowdrift game*. Doebeli *et al.* [235] provide a good description:

[T]wo drivers are trapped on either side of a snowdrift and have the option of staying in their cars or removing the snowdrift. Letting the opponent do all the work is the best option, but if both players refuse to shovel they can't get home. The essential feature of the snowdrift game is that

6.3 A Fokker-Planck Equation for Adaptive Dynamics

defection is better than cooperation if the opponent cooperates but worse if the opponent defects.

If the amount of cooperation is a continuously variable quantity, then this fits into the framework we have developed in this chapter. The payoff to an agent who cooperates an amount x when met with another agent who cooperates to the extent y is

$$A(x; y) = -C(x) + B(x + y). \quad (6.38)$$

Note that if we take a partial derivative with respect to x or with respect to y , we pick up a B' either way.

The mutation-selection equilibrium distribution for the Snowdrift game is

$$P^*(x) \propto \frac{N_e}{u(x)\nu} \exp \left(2N_e \int_0^x \frac{dx'}{A(x'; x')} \left[-C'(x') + \left(1 + \frac{\sigma - 1}{\sigma + 1} \right) B'(2x') \right] \right). \quad (6.39)$$

Again, simplifying the integral is not straightforward, but we can find the position where $P^*(x)$ is peaked.

Defining the convenient abbreviation

$$r = \frac{\sigma - 1}{\sigma + 1}, \quad (6.40)$$

the argument of the exponential is extremized where

$$\frac{B'(2x)}{C'(x)} = \frac{1}{1 + r}. \quad (6.41)$$

In the deterministic limit, the Snowdrift game is known to exhibit interesting behavior with the quadratic cost and benefit functions

$$C(x) = c_2 x^2 + c_1 x, \quad B(x) = b_2 x^2 + b_1 x. \quad (6.42)$$

Given these forms for the cost and benefit curves, then Eq. (6.41) is satisfied when

$$x = \frac{c_1 - (1 + r)b_1}{4(1 + r)b_2 - 2c_2}. \quad (6.43)$$

At $\sigma = 1$, we have $r = 0$, and our expression for x reduces to

$$x = \frac{c_1 - b_1}{4b_2 - 2c_2}, \quad (6.44)$$

which is the equilibrium value that Doebeli *et al.* derive for deterministic dynamics in a panmictic population [235].

The values of b_1 , b_2 , c_1 and c_2 govern whether or not the fixed point of the deterministic dynamics is stable. By adjusting these parameters appropriately, we can have

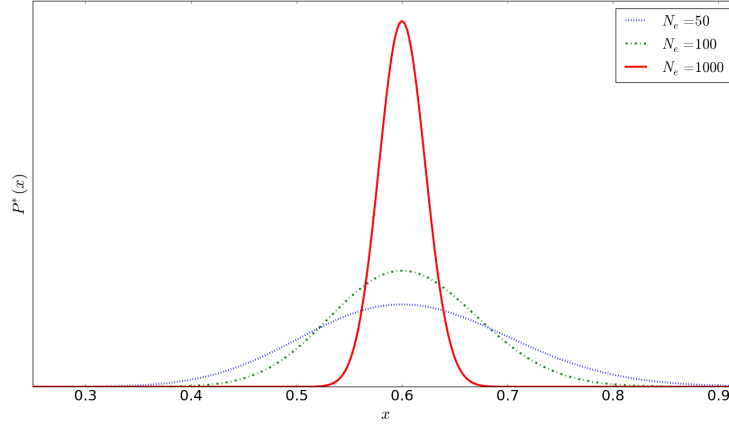


Figure 6.1: Steady-state probability distributions for the Snowdrift game, computed with three different choices of the effective population size N_e . As we increase N_e from 1 to 10 and then 100, the distribution becomes narrower. The peak of the distribution is, in each case, located at the ESS value of the deterministic dynamics, $x = 0.6$ ($b_1 = 7$, $b_2 = -1.5$, $c_1 = 4.6$, $c_2 = -1.0$, $\sigma = 1.0$).

fixed points at the same value of x but opposite stabilities. For example,

$$b_1 = 7, \quad b_2 = -1.5, \quad c_1 = 4.6, \quad c_2 = -1.0 \quad (6.45)$$

yields a fixed point at $x = 0.6$, as does

$$b_1 = 3.4, \quad b_2 = -0.5, \quad c_1 = 4.0, \quad c_2 = -1.5. \quad (6.46)$$

However, in the former case, $x = 0.6$ is an Evolutionary Stable Strategy, while in the latter, it is a repeller point [235]. The analogue of this in the stochastic case is the fact that two probability distributions can have extrema in the same positions, one having a minimum and the other a maximum in that location. Now, we investigate this issue in more detail. In particular, we'd like to know how varying the structure coefficient σ affects stability.

Whether an extremum of $P^*(x)$ is a minimum or a maximum depends on the second derivative of the integral inside the exponential, which is the first derivative of the integrand. Applying the quotient rule to the integrand yields a ratio whose denominator is $A(x; x)^2$. This is always positive, so the sign can only depend on the sign of the numerator. Furthermore, the numerator itself simplifies at an extremum, leaving us with the condition

$$2(1 + r)B''(2x) - C''(x) < 0. \quad (6.47)$$

For quadratic cost and benefit functions, this becomes

$$2(1+r)b_2 - c_2 < 0. \quad (6.48)$$

If this inequality is satisfied, then the extremum x is a maximum, and thus is evolutionarily stable.

Incorporating stochastic effects by way of our Fokker–Planck equation goes beyond prior work on the continuous Snowdrift game. We can go further by varying the structure coefficient σ , which corresponds to implementing the Snowdrift game with different short-term ecological dynamics. We plot typical results in Figure 6.2. Generally, as we increase σ , the peak of the steady-state distribution moves to higher x , indicating an increased disposition to cooperation.

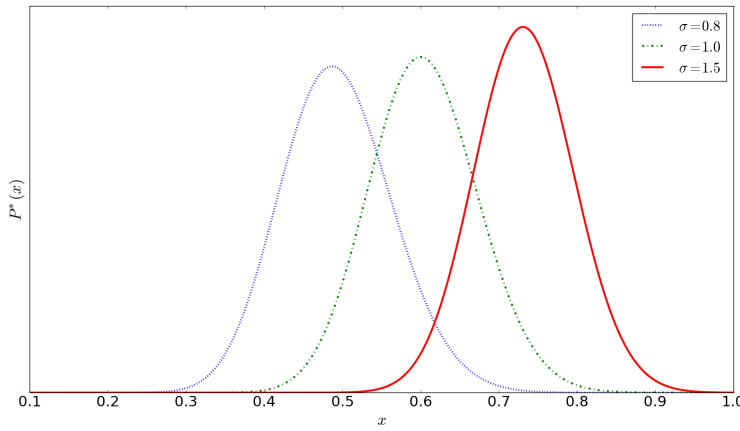


Figure 6.2: Steady-state probability distributions for the Snowdrift game, computed with three different choices of the Tarnita structure coefficient σ . As we increase σ , the peak of the steady-state distribution moves to higher x , indicating an increased disposition to cooperation ($b_1 = 7$, $b_2 = -1.5$, $c_1 = 4.6$, $c_2 = -1.0$, $N_e = 100$).

6.4 Concurrent Mutations, Discreteness and Multi-strategy Games

The message of adaptive dynamics, whether deterministic or stochastic, is that calculations become easier when traits vary continuously and mutations are both small and rare. In essence, adaptive dynamics illustrates the principle that when we can use Taylor expansion, we can simplify. Now, we consider factors that can spoil that clean conceptual simplicity.

6 Stochastic Adaptive Dynamics

Suppose that each player in a population can choose from the same set of n strategies, where n is possibly larger than two. Let's assume that the game interactions still occur on a pairwise basis. That is, we can write the payoffs as a matrix a_{ij} , where i and j both range from 1 to n . We can define four different meaningful averages:

$$\bar{a}_{**} = \frac{1}{n} \sum_{i=1}^n a_{ii}, \quad (6.49)$$

$$\bar{a}_{r*} = \frac{1}{n} \sum_{i=1}^n a_{ri}, \quad (6.50)$$

$$\bar{a}_{*r} = \frac{1}{n} \sum_{i=1}^n a_{ir}, \quad (6.51)$$

$$\bar{a} = \frac{1}{n^2} \sum_{i,j=1}^n a_{ij}. \quad (6.52)$$

If the effect of selection is not too strong, then the condition for strategy r to be evolutionarily favored is

$$\sigma_1(a_{rr} - \bar{a}_{**}) + \sigma_2(\bar{a}_{r*} - \bar{a}_{*r}) + \sigma_3(\bar{a}_{r*} - \bar{a}) > 0. \quad (6.53)$$

The coefficients σ_1 , σ_2 and σ_3 depend upon the population structure and the update rule, just as σ did for two-strategy games, but they are independent of the payoff function [236, 237]. By dividing through by a nonzero structure coefficient, we can convert Eq. (6.53) into a two-parameter condition.

This becomes relevant to adaptive dynamics if multiple mutations can be present in the population at the same time, even if only for brief intervals. This is, in fact, a biologically important scenario [144]. If an individual in the population can pick from among more than two strategies, then we will in general require at least two structure coefficients in order to decide which strategy will end up dominant. This rather spoils the convenience of having one σ that controls the outcome.

What if the trait value of interest is not continuously variable, but instead discretized? One problem which motivates an investigation into discretized adaptive dynamics is the evolution of drug resistance. Take *Plasmodium falciparum*, a parasitic protozoan that is responsible for malaria. The standard treatment for it was a drug called chloroquine, but over the years, *P. falciparum* evolved the ability to resist the medication [66, 238]. Chloroquine resistance (CQR) arose independently multiple times during the twentieth century—quite probably, more times than we know about, because any mutation that arises in the wild must proliferate to an extent before researchers can detect it. A recent study [239] examined the evolution of CQR in detail and found that there exist multiple paths of mutations leading from low resistance to high. Intermediate stages provide partial resistance, the levels of which can be quantified.

Drug resistance can be treated in terms of evolutionary games, particularly when

some organisms produce compounds that are harmful to others [240]. Furthermore, we expect the same general pattern—a spread of intermediates between extreme trait values, discretized fundamentally by the genetic code—in other microbial social behaviors.

This suggests the following idealized scenario:

Most of the time, the population is genetically uniform. Occasionally, a mutation arises, and the new variety challenges the established resident type for dominance. We again make the approximation that at most one mutant type is ever present at any given time, but unlike before, we do not treat the trait value or the mutation step size as continuously variable. Mutations are rare, but they are no longer small.

In a two-player game with two strategies, a resident strategy x is favored over a mutant strategy x' provided that

$$\sigma[A(x; x) - A(x'; x')] + A(x; x') - A(x'; x) > 0. \quad (6.54)$$

Let the separation between the possible values of x be Δ . This defines the mutation step size. If a strategy x is favored over both $x + \Delta$ and $x - \Delta$, then we must have

$$\begin{aligned} \sigma[A(x; x) - A(x + \Delta; x + \Delta)] + A(x; x + \Delta) - A(x + \Delta; x) &> 0, \\ \sigma[A(x; x) - A(x - \Delta; x - \Delta)] + A(x; x - \Delta) - A(x - \Delta; x) &> 0. \end{aligned} \quad (6.55)$$

A value of x where these conditions are met is a local equilibrium.

Because the mutational steps are discrete, we cannot trust the decomposition of a multiplayer game into dyadic ones. Therefore, if the basic game is a multiplayer interaction, then we need a criterion analogous to Eq. (6.54) for K -player games with two strategies.

Denote the strategies by A and B . Let a_j and b_j be the payoffs to a focal A - or B -type player, respectively, given that they interact with j neighbors who play strategy A . Assuming weak selection, strategy A is favored if

$$\sum_{j=0}^{K-1} \sigma_j (a_j - b_{K-1-j}) > 0. \quad (6.56)$$

That is, for two-strategy games, the number of structure coefficients grows linearly with the number of players [237, 241]. This means that when the trait value x is discretized, we will need more than one structure coefficient to locate the equilibrium point, even if mutations are rare and the population is monomorphic in between competition events.

A more general and potentially more realistic modeling method is to picture the allowed trait values not as points on a line, but as vertices in a graph. Each vertex represents a possible genotype, for example, one of the variations of the *PfCRT* gene responsible for chloroquine resistance. We connect two vertices with an edge if a mutation, like a substitution of nucleotides, can convert between those two forms of the gene. If we treat the competition between two types as an evolutionary game, we can use the payoff function and the structure coefficients to assign a direction to each edge of the graph. Local equilibria are sinks of the flow, *i.e.*, vertices that

have all their edges pointing inwards. (More specifically, such points should “collect probability,” even though it is not *guaranteed* that the system must always move along the direction of the edges. The σ -rule analysis tells us which fixation probability is greater, not the probabilities themselves. Moving against an arrow is possible, but we bet against its happening. This is another place where the discrete case is more complicated than the continuous. In the adaptive dynamics of continuous quantities, the smoothness of all the functions involved means that we only need to know which fixation probability is greater, not their specific values [230].) Assuming that none of the mutations fundamentally change the population structure and life cycle, then we can use the same structure coefficients for all the edges. This may not always be a reasonable assumption. For example, a mutation in a virus could allow it to jump into a new host species. In that event, using the same σ values before and after the mutation would clearly not be reasonable. However, for mutations of lesser consequence, it seems a viable starting point.

Up until this point, we have treated the mutation rate $u(x)$ as independent of the trait value x . But we can argue that this does not have to be the case. We saw back in Chapter 2 that many different nucleotide sequences can map to the same protein. Because these sequences are equivalent as far as their products are concerned, within the space of all possible sequences there exist regions inside of which movement is unaffected by selection pressure.

Mutations happen stochastically with a certain probability per base pair. With more genetic material overall, we expect more mutations. The rate at which we see changes in the trait value x depends upon how easily a mutation can push a genetic sequence from the space that maps to one phenotype into a space that maps to another. The volumes and surface areas of these spaces do not have to be constant over x ! Consequently, some values of x can be more *robust* than others. We can define the robustness of a nucleotide sequence quantitatively: we consider the set of all sequences that can be produced from the original by mutations, and we find the fraction of that set for which the phenotype is unchanged. Then, we can average this quantity over all sequences that map to the same phenotype in order to define a robustness for that phenotype [242]. For systems that are small enough to study with simulations—the folding of short RNA sequences is a popular choice—values of the robustness can vary over an order of magnitude from one phenotype to another [243]. In order to apply adaptive dynamics, we originally assumed that the trait value x could vary smoothly. A heritable trait with an effectively continuous spectrum of possible values will almost certainly be the product of many individual genes and regulatory elements, acting in concert. The number of nucleotide sequences that will map to the same value of x will, therefore, be vast. Nevertheless, the same considerations should apply. If the robustness varies with x , then so too will the effective mutation rate $u(x)$. The more rapid the variations in $u(x)$, the more the effects will be seen in the mutation–selection equilibrium distribution $P^*(x)$, particularly for smaller population sizes.

We have discussed the cases of discrete and of continuous x . Somewhat of a conceptual intermediary is the case where x is continuous, but *mutation step sizes can be large*. When we derived the Fokker–Planck equation (6.9), we assumed explicitly that changes are dominated by small increments, and larger fluctuations are all negligible.

Then we carried this assumption over into adaptive dynamics, where we said that the mutation step size distribution $\mu(z)$ was sufficiently narrow that we only needed to know its variance. However, considerations of *systems biology* suggest that in some situations, we ought to allow $\mu(z)$ to be much wider. The interaction networks of intracellular components often have long-tailed degree distributions: most components have few interaction partners, but a few have many, and there is a continuum of variation in between the extremes [244]. This hints that many mutations will have limited effects, while a small number—those that directly impact the components of highest network degree—will sway the trait value x much more strongly.

The distribution of mutation effect sizes is currently a research topic with a great many question marks, both empirically and theoretically [245]. Furthermore, it is not clear how applicable the extant theoretical results are to our problem; their basic assumptions were not formulated with adaptive dynamics in mind. We therefore have a certain freedom to go off in new directions. One interesting possibility comes from the study of *Lévy flights*, random walks in which the distribution of step lengths is long-tailed. Investigating this type of stochastic process reveals that the Fokker–Planck equation must be replaced with an analogue that is written using *fractional derivatives*.

Again, this is a situation in which we should not expect that we could decompose a many-player game into dyads.

6.5 Interspecies Interactions

We now return to the continuous Prisoner’s Dilemma. Pleasingly, the peak value of x computed by Eq. (6.37) is also the equilibrium value for the deterministic time-evolution equation derived by Allen, Nowak and Dieckmann [230]. Up to a constant prefactor of little interest,

$$\frac{dx}{dt} \propto -C'(x) + \frac{\sigma - 1}{\sigma + 1} B'(x). \quad (6.57)$$

What must be satisfied for an equilibrium point of this dynamical rule to be stable? Let $dx/dt = 0$ at $x = \bar{x}$, and define the separation from equilibrium by $\xi = x - \bar{x}$. Rescale time as convenient, so that the proportionality becomes an equality. Then

$$\frac{d\xi}{dt} = \frac{d}{dt}(x - \bar{x}) \quad (6.58)$$

$$= -C'(\bar{x} + \xi) + \frac{\sigma - 1}{\sigma + 1} B'(\bar{x} + \xi) \quad (6.59)$$

$$\approx -[C'(\bar{x}) + \xi C''(\bar{x})] + \frac{\sigma - 1}{\sigma + 1} [B'(\bar{x}) + \xi B''(\bar{x})] \quad (6.60)$$

$$= \xi \left[-C''(\bar{x}) + \frac{\sigma - 1}{\sigma + 1} B''(\bar{x}) \right]. \quad (6.61)$$

6 Stochastic Adaptive Dynamics

This is a linear response equation for the displacement ξ :

$$\frac{d\xi}{dt} = -\xi \left[C''(\bar{x}) - \frac{\sigma - 1}{\sigma + 1} B''(\bar{x}) \right]. \quad (6.62)$$

Since we required that both $B(x)$ and $C(x)$ be twice differentiable, this is fine. The condition that the equilibrium $\bar{x}$ be stable is that the quantity in square brackets is positive. Therefore, stability requires

$$C''(\bar{x}) > \frac{\sigma - 1}{\sigma + 1} B''(\bar{x}). \quad (6.63)$$

Now, let us consider what happens if we have *multiple* species, each evolving in accord with the deterministic dynamics specified by a cost function and a benefit function. For $i = 1, 2, \dots, M$, we have, if the species do not interact,

$$\frac{dx_i}{dt} \propto -C'_i(x_i) + \frac{\sigma_i - 1}{\sigma_i + 1} B'_i(x_i). \quad (6.64)$$

We assume that a stable equilibrium $\bar{x}_i$ exists for each species. So, in terms of the displacements $\xi_i = x_i - \bar{x}_i$,

$$\frac{d\xi_i}{dt} = -\xi_i \left[C''_i(\bar{x}_i) - \frac{\sigma_i - 1}{\sigma_i + 1} B''_i(\bar{x}_i) \right]. \quad (6.65)$$

What happens if the species in this ecosystem start to interact with one another? The most straightforward way to incorporate interactions into this dynamical system is to include cross terms:

$$\frac{d\xi_i}{dt} = -\xi_i \left[C''_i(\bar{x}_i) - \frac{\sigma_i - 1}{\sigma_i + 1} B''_i(\bar{x}_i) \right] + \alpha \sum_{j=1}^M J_{ij} \xi_j. \quad (6.66)$$

The parameter α controls the strength of the interactions, and the matrix element J_{ij} indicates how much species j influences species i .

We lose nothing of real interest if we assume that the factors multiplying the ξ_i are equal for all i , and so we can set them to unity. This is a further simplification, but a reasonable one, since we expect that the timescales with which each species in isolation settles to its own equilibrium are roughly equal.

The equilibrium we had without interactions was stable. It will remain stable with the interactions turned on if the eigenvalues λ_a of the coupling matrix J all satisfy the inequality

$$\alpha \lambda_a - 1 \leq 0, \quad (6.67)$$

which is the same thing as saying that the largest eigenvalue is less than or equal to $1/\alpha$. Consequently, if the elements of the matrix J are chosen at *random*, then the *probability* that the equilibrium will remain stable is the probability that the largest eigenvalue of J is bounded by $1/\alpha$.

6.5 Interspecies Interactions

This is exactly the kind of problem studied in random matrix theory. In fact, we have recovered (with a new meaning for the symbols) an old problem of May [246], which has deep connections to the statistical physics of coupled random variables [247].

7 Spatial Stochastic Mechanics

The goal of this chapter is to provide, by analytical means, quantitative predictions for the values of some observations made in Chapters 3 and 4. We will do this not by solving the specific models we studied there, but rather by understanding related models, and seeing how results gleaned from those solutions can be translated over to our original problems of interest. We touched upon this method of exploration in Chapter 3, when we introduced the concept of a universality class (§3.3.6). We learn about one system in a universality class by studying another and finding the features which apply across the class in its entirety. Now, it is time to apply this method in detail. We will begin by grounding ourselves in the essential ideas of *renormalization* theory, which underpins the study of universality classes. Then, we will build the infrastructure necessary to express the stochastic spatial models of Chapters 3 and 4 analytically. With that development complete, we will be able to justify the theoretical predictions to which we compared the phase-transition phenomena of those models.

Our subject for the next few sections will be “the renormalization group.” This term is a little like the Holy Roman Empire, in that the latter, as Voltaire observed, is neither holy, Roman nor an empire. The word *group* in this context is a technical term, but mathematically speaking, it is not really the *right* technical term. *Renormalization* is a bit of jargon that hails from quantum electrodynamics, and it dates to a time many years before people figured out what they were doing and why their techniques actually worked. Consequently, it is not very indicative. Even *the* isn’t quite right: it carries the implication that “*the* renormalization group” is a single, specific construction, like “the quadratic formula.”

Since the full name of the subject is a misnomer, it is convenient to elide that name as much as possible, and so we will use the common abbreviation *RG*. We begin, therefore, our discussion of *RG theory* and *RG transformations*.

7.1 The Central Limit Theorem by RG Transformations

To ease ourselves into RG theory, we will revisit a topic we addressed before, this time approaching it from a different, albeit related perspective. Our first example of RG theory is the central limit theorem, which we first proved in §5.7. A loose but useful statement of this result is that the sum total of many uncorrelated small influences is statistically distributed following a Gaussian curve.

When we introduced complex systems, back in Chapter 2, we defined systems made of components, and we quantified relationships among those components using information theory. Let us bring the central limit theorem into this context. Consider a system with many components, each of which is a random variable. We describe the

7 Spatial Stochastic Mechanics

system probabilistically by ascribing a joint distribution to the set of random variables. In this case, we postulate that all higher-order complexities vanish, meaning that the probability distribution factors neatly into single-variable distributions:

$$p(X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots) = p(X_1 = x_1)p(X_2 = x_2)p(X_3 = x_3) \cdots \quad (7.1)$$

To start with, let the components of our system be coins, which we say are independent of each other and equally likely to come up either heads (+1) or tails (−1). A snapshot of this system will be, we expect, a speckling of pluses and minuses in roughly equal proportions (we might get a large fluctuation one way or the other, but we're not gambling on that). What happens if we blur our view, grouping together pairs of components?

The total value of a pair of coins can be −1, 0 or +1. For a single coin, we have

$$p_1(-1) = \frac{1}{2}, \quad p_1(+1) = \frac{1}{2}. \quad (7.2)$$

There are two ways the total could sum to 0, and so

$$p_2(-2) = \frac{1}{4}, \quad p_2(0) = \frac{1}{2}, \quad p_2(+2) = \frac{1}{4}. \quad (7.3)$$

We see that the distribution stays symmetrical and becomes peaked in the middle.

Following the same logic, we can repeat the blurring procedure. The arithmetic becomes more involved, but the concept is straightforward. After just a few repetitions, the symmetrical, peaked distribution comes to resemble a Gaussian curve, as shown in Figure 7.1.

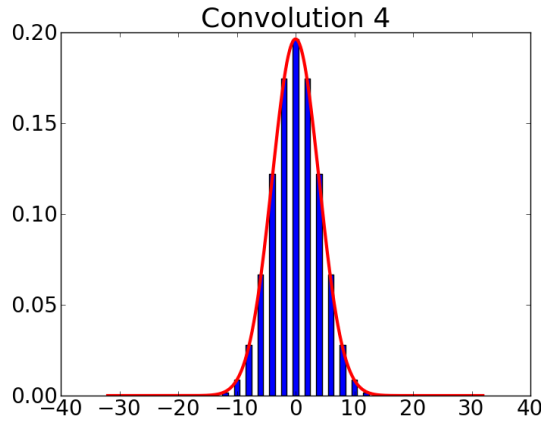


Figure 7.1: Discrete probability distribution obtained by coarse-graining uncorrelated fair coins four successive times. The solid curve is a Gaussian with mean zero and variance $\sigma^2 = 16$.

7.1 The Central Limit Theorem by RG Transformations

What happens if we start our gambling on the assumption that the coins are biased? Say, for example, that we had chosen

$$p_1(-1) = 0.1, \quad p_1(+1) = 0.9. \quad (7.4)$$

Then we would have

$$p_2(-2) = 0.01, \quad p_2(0) = 0.18, \quad p_2(+2) = 0.81. \quad (7.5)$$

This p_2 is not symmetrical at all! However, as we repeat the blurring procedure, the asymmetry partially washes out. Examining Figure 7.2, we see that the statistics for the coarse-grained view are still approximately Gaussian, but the mean of the Gaussian curve has shifted in the direction of the original tilt.

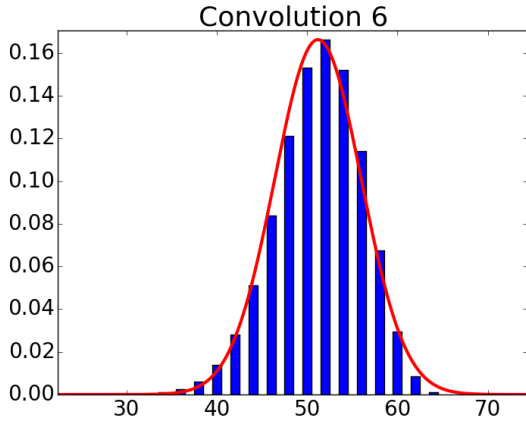


Figure 7.2: Discrete probability distribution obtained by coarse-graining uncorrelated loaded coins six successive times. The solid curve is a Gaussian with mean 51.2 and variance $\sigma^2 = 23.0$.

Let's explore the effects of successive coarse-graining in more generality. Take $p(x)$ to be the probability distribution for each of our original variables, and suppose for convenience that its mean is zero. When we convolve $p(x)$ with itself, we double all its cumulants. Now, after we coarse-grain, let us *rescale the axis* by a factor $1/\sqrt{2}$. This restores the variance to its original value, because the variance involves two powers of x . We repeat the operations of coarse-graining and rescaling K times in succession. The mean and the variance remain unchanged, but what about the higher cumulants?

The cumulants of the original distribution are $\langle x^n \rangle_c$. Each coarse-graining doubles $\langle x^n \rangle_c$, and each rescaling multiplies $\langle x^n \rangle_c$ by a factor $(1/\sqrt{2})^n$. So, if $\langle z_K^n \rangle_c$ denotes

7 Spatial Stochastic Mechanics

the result of K iterations, then

$$\langle z_K^n \rangle_c = \langle x^n \rangle_c \left(\frac{2}{2^{n/2}} \right)^K = \langle x^n \rangle_c \left(2^{1-n/2} \right)^K. \quad (7.6)$$

For $n > 2$, repeating the operations of coarse-graining and rescaling will send $\langle z_K^n \rangle_c$ to zero. This is just the conclusion we drew in §5.7, phrased in terms of an iterative procedure. We say that the higher cumulants are *irrelevant*, because they shrink under the iteration.

If we hadn't invested time in developing our understanding of cumulants, we could have arrived at the same conclusion by going back to the basics of Fourier transforms. (This is the approach taken, for example, in Sethna's textbook [248].) We define the Fourier transform of a probability density function $p(x)$ as

$$\mathcal{F}[p](k) = \int_{-\infty}^{\infty} dx e^{-ikx} p(x) = \tilde{p}(k). \quad (7.7)$$

The inverse transformation is

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ikx} \tilde{p}(k). \quad (7.8)$$

We coarse-grain by convolution:

$$(p \star p)(x) = \int_{-\infty}^{\infty} dy p(x-y)p(y). \quad (7.9)$$

By the convolution theorem, the Fourier transform of a convolution is the product of the Fourier transforms of the functions convolved. In this case, convolving a curve with itself,

$$\mathcal{F}[p \star p](k) = (\tilde{p}(k))^2. \quad (7.10)$$

The next step is to re-scale the curve, under the assumption that the mean is zero. This has the effect of creating a new function,

$$\mathcal{S}_{\sqrt{2}}[p](x) = \sqrt{2}p(\sqrt{2}x). \quad (7.11)$$

The prefactor of $\sqrt{2}$ is necessary to preserve normalization, as we can see by integrating:

$$\begin{aligned} \int_{-\infty}^{\infty} dx \mathcal{S}_{\sqrt{2}}[p](x) &= \sqrt{2} \int_{-\infty}^{\infty} dx p(\sqrt{2}x) \\ &= \sqrt{2} \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2}} p(y) \\ &= 1. \end{aligned} \quad (7.12)$$

7.1 The Central Limit Theorem by RG Transformations

The change of length scale is equivalent to an inverse change of frequency:

$$\begin{aligned}\mathcal{F}[\mathcal{S}_{\sqrt{2}}[p]](k) &= \int_{-\infty}^{\infty} dx e^{-ikx} \sqrt{2} p(\sqrt{2}x) \\ &= \sqrt{2} \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2}} e^{-iky/\sqrt{2}} p(y) \\ &= \tilde{p}(k/\sqrt{2}).\end{aligned}\tag{7.13}$$

It follows that the Gaussian curve

$$p^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)\tag{7.14}$$

is a fixed point of the coarse-grain-and-rescale transformation. We shall now prove this explicitly.

For brevity, we denote this composite transformation by $\mathcal{T}$. The $\mathcal{T}$ operator effects a convolution and a rescaling:

$$\mathcal{T}[p](x) = \mathcal{S}_{\sqrt{2}}[p \star p](x) = \sqrt{2} \int_{-\infty}^{\infty} dy p(\sqrt{2}x - y) p(y).\tag{7.15}$$

Explicitly,

$$\mathcal{T}[p^*](x) = \sqrt{2} \int_{-\infty}^{\infty} dy \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(\sqrt{2}x - y)^2}{2\sigma^2}\right) \exp\left(-\frac{y^2}{2\sigma^2}\right).\tag{7.16}$$

Expanding out and combining the arguments of the exponentials,

$$\mathcal{T}[p^*](x) = \frac{\sqrt{2}}{2\pi\sigma^2} \int_{-\infty}^{\infty} dy \exp\left(-\frac{2x^2 - 2\sqrt{2}xy + 2y^2}{2\sigma^2}\right).\tag{7.17}$$

With a bit of algebra,

$$\begin{aligned}2x^2 - 2\sqrt{2}xy + 2y^2 &= 2y^2 - 2(\sqrt{2}y)x + 2x^2 \\ &= 2y^2 - 2(\sqrt{2}y)x + x^2 + x^2 \\ &= (\sqrt{2}y - x)^2 + x^2,\end{aligned}\tag{7.18}$$

we can make the integral over y a Gaussian integral, which we know how to evaluate. This eliminates the y dependence, leaving us with

$$\mathcal{T}[p^*](x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) = p^*(x),\tag{7.19}$$

as desired.

7 Spatial Stochastic Mechanics

The Fourier transform of p^* is

$$\tilde{p}^*(k) = \mathcal{F}[p^*] = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{-ikx - x^2/(2\sigma^2)} = \exp\left(-\frac{k^2\sigma^2}{2}\right). \quad (7.20)$$

Generally, the Fourier transform of a $\mathcal{T}$ -transformed function is

$$\tilde{\mathcal{T}}[\tilde{p}](k) = \mathcal{F}[\mathcal{T}[p]](k) = \left(\tilde{p}(k/\sqrt{2})\right)^2, \quad (7.21)$$

with which it is easy to verify that

$$\tilde{\mathcal{T}}[\tilde{p}^*](k) = \mathcal{F}[\mathcal{T}[p^*]](k) = \tilde{p}(k). \quad (7.22)$$

That is, p^* is a fixed point in the space of probability distributions, and $\tilde{p}^*$ is a fixed point of the analogous transformation in the space of Fourier representations.

Let p be a distribution which is close to p^* :

$$p(x) = p^*(x) + \epsilon f(x). \quad (7.23)$$

Then, because the Fourier transform is linear,

$$\tilde{p}(k) = \tilde{p}^*(k) + \epsilon \tilde{f}(k). \quad (7.24)$$

What are the eigenfunctions and eigenvalues of $\mathcal{T}$?

$$\mathcal{T}[p^* + \epsilon f_n] = p^* + \lambda_n \epsilon f_n + \mathcal{O}(\epsilon^2), \quad (7.25)$$

$$\tilde{\mathcal{T}}[\tilde{p}^* + \epsilon \tilde{f}_n] = \tilde{p}^* + \lambda_n \epsilon \tilde{f}_n + \mathcal{O}(\epsilon^2). \quad (7.26)$$

Using Eq. (7.21),

$$\tilde{\mathcal{T}}[\tilde{p}^* + \epsilon \tilde{f}_n] = \left[\tilde{p}^*(k/\sqrt{2}) + \epsilon \tilde{f}_n(k/\sqrt{2})\right]^2 \quad (7.27)$$

$$= \left[\tilde{p}^*(k/\sqrt{2})\right]^2 + 2\epsilon \tilde{p}^*(k/\sqrt{2}) \tilde{f}_n(k/\sqrt{2}) + \mathcal{O}(\epsilon^2) \quad (7.28)$$

$$= \tilde{p}^*(k) + 2\epsilon \tilde{p}^*(k/\sqrt{2}) \tilde{f}_n(k/\sqrt{2}) + \mathcal{O}(\epsilon^2). \quad (7.29)$$

Therefore,

$$\tilde{f}_n(k) = \frac{2}{\lambda_n} \tilde{p}^*\left(\frac{k}{\sqrt{2}}\right) \tilde{f}_n\left(\frac{k}{\sqrt{2}}\right). \quad (7.30)$$

Proposal: let

$$\tilde{f}_n(k) = (ik)^n \tilde{p}^*(k). \quad (7.31)$$

7.1 The Central Limit Theorem by RG Transformations

With this ansatz,

$$\tilde{p}^* \left(\frac{k}{\sqrt{2}} \right) \tilde{f}_n \left(\frac{k}{\sqrt{2}} \right) = \left(\frac{ik}{\sqrt{2}} \right)^n \tilde{p}^* \left(\frac{k}{\sqrt{2}} \right) \quad (7.32)$$

$$= \frac{(ik)^n}{(\sqrt{2})^n} \tilde{p}^*(k) \quad (7.33)$$

$$= \frac{1}{(\sqrt{2})^n} \tilde{f}_n(k). \quad (7.34)$$

Consequently,

$$\tilde{f}_n(k) = (\sqrt{2})^n \tilde{p}^* \left(\frac{k}{\sqrt{2}} \right) \tilde{f}_n \left(\frac{k}{\sqrt{2}} \right), \quad (7.35)$$

meaning that the eigenvalue λ_n is given by

$$\frac{2}{\lambda_n} = (\sqrt{2})^n = 2^{n/2} \Rightarrow \lambda_n = 2^{1-n/2}. \quad (7.36)$$

The first eigenmode, associated with the eigenvalue $\lambda_0 = 2$, is

$$\tilde{f}_0(k) = \tilde{p}^*(k). \quad (7.37)$$

When applied as a perturbation to $\tilde{p}^*$, this yields

$$\tilde{p}(k) = \tilde{p}^*(k) + \epsilon \tilde{p}^*(k), \quad (7.38)$$

meaning that

$$p(x) = (1 + \epsilon) p^*(x). \quad (7.39)$$

This is not a normalized probability distribution function, unless $\epsilon = 0$. Therefore, we can neglect the λ_0 mode as meaningless.

The next eigenfunction is

$$\tilde{f}_1(k) = ik \tilde{p}^*(k). \quad (7.40)$$

Applying this perturbation to $\tilde{p}^*$ yields a probability distribution

$$p(x) = \frac{1}{2\pi} \int dk (1 + \epsilon ik) e^{ikx} \tilde{p}^*(k). \quad (7.41)$$

Because ϵ is a small number, we can use a Taylor approximation:

$$p(x) = \frac{1}{2\pi} \int dk e^{ik\epsilon} e^{ikx} \tilde{p}^*(k) = \frac{1}{2\pi} \int dk e^{ik(\epsilon+x)} \tilde{p}^*(k). \quad (7.42)$$

Therefore,

$$p(x) = p^*(x + \epsilon). \quad (7.43)$$

That is, perturbing in the λ_1 eigenmode is equivalent to shifting the mean of $p^*(x)$.

The operation of coarse-graining and then rescaling is our first example of an RG

transform, and the cumulants provide our first encounter with the concept that parameters can be relevant or irrelevant under an RG transform. We think of a *flow* in the space of parameters: in this case, the flow induced by the RG transform is the change in the cumulant values. The flow takes us to a *fixed point* which is, in this example, a Gaussian curve. Generally, a quantity is *relevant* if repeated applications of the RG transform cause it to grow, and a quantity is *irrelevant* if it shrinks as we approach the RG fixed point. (Quantities can also be *marginal*, if a linear stability analysis of an RG fixed point fails to indicate whether they will increase or decrease.)

This example also illustrates how fixed points of RG flows relate to *universality*. Gaussian curves show up throughout the sciences, because they arise whenever one considers the sum total effect due to a large number of uncorrelated small influences. Furthermore, once you understand one Gaussian curve, you've pretty much understood them all. This is the essence of a universality class, as we defined that concept in Chapter 3. It is the fact that so many details are irrelevant, in the RG sense, that makes the Central Limit Theorem so powerful.

Using the calculus of variations, one can show that Gaussian curves maximize the Shannon index (also known as the Shannon entropy) for a given variance. That is, if we fix the variance to be σ^2 , then the largest value of the Shannon index compatible with this constraint is

$$S_G(\sigma^2) = \frac{1}{2} \log_2(2\pi e\sigma^2). \quad (7.44)$$

As we repeat the RG transform, the variance remains constant, and the probability distribution looks more and more Gaussian, meaning that the Shannon index will approach S_G from below. The Shannon index is an example of a quantity which increases along with the RG flow, attaining a maximum at the flow's fixed point. When RG ideas are applied in the realm of field theory, this idea becomes the *c*-theorem of Zamolodchikov [249].

Finally, we saw that *exponents* turn out to be important quantities: in this example, they are the eigenvalues $\lambda_n = 2^{1-n/2}$. The importance of exponents that govern scaling behavior holds true across RG theory.

7.2 Isotropic Percolation

In the previous section, we did not put any spatial structure on the set of system components. When we decided to coarse-grain two variables together, any two variables were as good as any other pair. Now, we will move on to a problem where we can apply RG theory to a system that has spatial structure. Our next example of an RG analysis, which will move us closer to eco-evolutionary models, is the topic of *isotropic percolation*.

In Chapter 3, we introduced percolation as the study of flow through randomized media. Our focus was on *directed* percolation, in which there is a preferred direction (*e.g.*, downhill). It turns out to be somewhat easier to obtain quantitative results by relaxing this restriction and considering *isotropic* percolation, in which no spatial orientation is picked out as special.

Typically, the first step in defining an isotropic percolation problem is to construct a regular lattice. Each point, or vertex, in the lattice is connected to the same number of other vertices. One way to proceed is to imagine that each edge can be colored with one of two possible pigments, *e.g.*, white and black. The color of each edge is picked at random, independently of the choices for all other edges. For example, we can say that the probability an edge is marked with black is p , and so an edge chosen at random will be white with probability $1 - p$. We can then ask, for any value of p , what the sizes of the clusters formed by black edges will be. This defines a *bond percolation* problem.

Alternatively, we can make the lattice points our variables of interest. For example, we can say that each lattice point is either *filled* or *empty*, and we fill sites at random with probability p . Depending on the value of p , we will have different statistical distributions for the sizes of the clusters formed by adjacent filled points. This variant is known as *site percolation*.

What constitutes an RG transform in this context? It is easy to imagine what coarse-graining a lattice might mean: we can simply blur out the fine-scale details. This operation will create a new lattice based on the configuration of the original. One way to do this for site percolation is to consider each lattice point together with its immediate neighbors. We “collapse” this set of points down to a single site in the new lattice, and we choose whether the new site is filled or not based on how many of the original sites are filled. Applying this to all the sites in the original lattice creates a new graph in which the small-scale information has been blurred away. It is then convenient to “step back from the picture,” rescaling the lengths of the lattice edges so that neighboring points in the derived lattice are separated by the same distance as those in the original. This combination of coarse-graining followed by rescaling is an RG transform.

Suppose that p is small. Then the lattice sites will mostly be empty, though we will have some small islands of filled sites amidst the sea of emptiness. Applying the RG transform will tend to make each of these islands smaller, because filled sites on the edges of the islands will have too many empty neighbors for their coarse-grained images to be filled. So, heuristically speaking, the RG transform will create a new lattice configuration that looks like what we would find *for a smaller value of p* .

At the opposite extreme, suppose that p is close to 1. Then most of the sites will be filled, with some low density of empty regions. The RG transform will blur away small details, making the smallest empty regions vanish. Empty regions which aren’t quite small enough to vanish entirely will be shrunk in the transformed image. So, the result will be a lattice configuration that looks like what we would typically find *for a larger initial choice of p* .

If small values of the filling probability flow to even smaller ones, and large values flow to even larger ones, then there must be a “continental divide” somewhere in the middle, where p is mapped to *itself*. At this point, we expect to find clusters in a continuous range of sizes: the RG transform erases the smallest clusters, so there must be slightly larger ones to replace them, and so forth. When p equals this *critical* value p_c , the distribution of cluster sizes will be *scale-invariant*. Because the lattice will contain clumps of all sizes, when we study the overall properties of the system, the

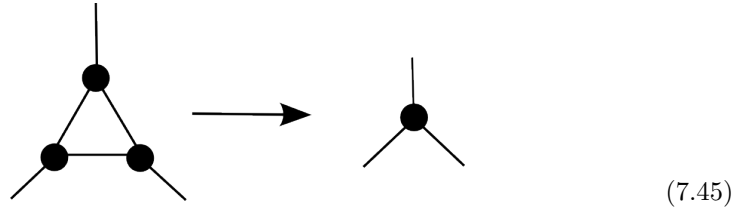
7 Spatial Stochastic Mechanics

details of how the connections between vertices look at the smallest scale should not matter much. This is the root of why we can group these systems into universality classes.

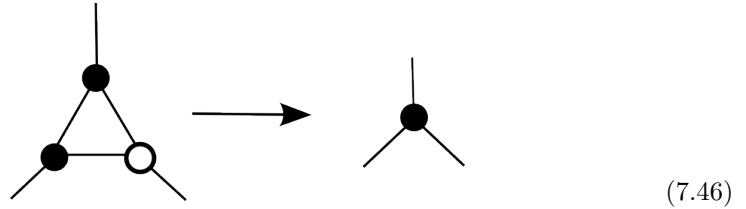
To make these qualitative considerations concrete, we will work through an example of site percolation. We take a triangular lattice and fill in the lattice sites at random. Let p denote the probability that a lattice site chosen at random is filled. We will see how a simple RG transform can send one value of p to another.

The coarse-graining operation should preserve the connectedness of clusters. This is rather tricky to enforce with a simple mapping, but one way of coming close is to use a *majority rule*. We map a triplet to a filled site if at least two of the original three sites are filled. This turns out to provide a simple approximation that gives remarkably good agreement with the best known results for the triangular lattice.

How does the majority rule create a mapping between values of p ? With probability p^3 , all three vertices of a triangle will be filled, and when this configuration occurs, we coarse-grain it to a single occupied site:

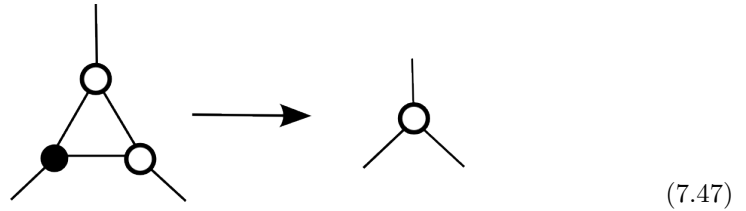


Likewise, if two out of three vertices are occupied, the corresponding site in the coarse-grained lattice is filled:

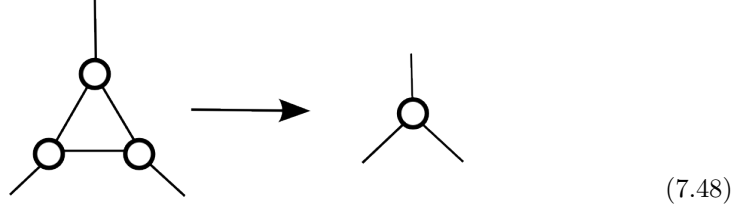


Because there are three ways to choose which vertex is unoccupied, this situation carries a symmetry factor, occurring with probability $3p^2(1-p)$.

We obtain an empty vertex in the coarse-grained lattice if two sites are open:



Or if all three vertices in the original triangle are unfilled:



Let p' denote the probability that a randomly chosen site in the coarse-grained lattice is filled. From the relations above, we see that

$$p' = p^3 + 3p^2(1 - p). \quad (7.49)$$

The fixed points of this mapping are given by

$$p_c = p_c^3 + 3p_c^2(1 - p_c). \quad (7.50)$$

Trivially, this equation is solved by 0 and by 1, and it has a less obvious solution at

$$\boxed{p_c = \frac{1}{2}}. \quad (7.51)$$

We will call this the *critical* value of the site-filling probability. What happens if p is almost but not quite equal to the critical value? Let us define

$$p = p_c + \delta p, \quad p' = p_c + \delta p'. \quad (7.52)$$

We can relate p and p' by the RG flow equation (7.49):

$$p_c + \delta p' = (p_c + \delta p)^3 + 3(p_c + \delta p)^2(1 - p_c - \delta p). \quad (7.53)$$

Expanding out the binomials, multiplying and simplifying, we arrive at

$$p_c + \delta p' = p_c + 6p_c(1 - p_c)\delta p. \quad (7.54)$$

Therefore,

$$\boxed{\delta p' = \frac{3}{2}\delta p}. \quad (7.55)$$

We see that a small deviation from the critical value becomes a larger one.

Are there exponents for the percolation problem as there were in the probability scenario we studied in the previous section? In fact, there are. The most important in practice is ν , which relates the characteristic length scale—the typical size of connected clumps—to the distance from the critical threshold:

$$\xi \propto |p - p_c|^{-\nu}. \quad (7.56)$$

7 Spatial Stochastic Mechanics

Here, we are measuring length in units of the lattice spacing. That the characteristic length ξ varies in this fashion is perhaps most directly appreciated by numerical simulation. We expect that the characteristic length will diverge at the critical threshold, thanks to the scale-invariance we described above.

Coarse-graining a lattice will produce a new grid in which the vertices are further apart. It is traditional to denote the factor by which the vertex separation is increased by the letter b . If $\xi(p)$ is the characteristic cluster size for filling probability p , then $\xi(p')$ is the characteristic cluster size on the lattice produced by coarse-graining, measured in units of the coarse-grained lattice spacing. But this must be related to the original average cluster size by the scaling factor:

$$\xi(p) = b\xi(p'). \quad (7.57)$$

For example, if we projected our original lattice on a wall and measured the typical cluster size to be four centimeters, then coarse-graining will produce a new image, containing fewer pixels per unit area, in which the typical cluster size must still be four centimeters.

If we rewrite our previous expression in terms of the distance from the threshold,

$$\xi(p_c + \delta p) = b\xi(p_c + \delta p'), \quad (7.58)$$

substituting in the scaling form involving the exponent ν yields

$$|\delta p|^{-\nu} = b|\delta p'|^{-\nu}. \quad (7.59)$$

We know from Eq. (7.55) how δp and $\delta p'$ are related, and so we can say that

$$|\delta p|^{-\nu} = b \left(\frac{3}{2} \right)^{-\nu} |\delta p|^{-\nu}. \quad (7.60)$$

Canceling the common factor and solving for ν ,

$$\nu = \frac{\log b}{\log(3/2)}. \quad (7.61)$$

We can show geometrically that for this coarse-graining operation on the triangular lattice, $b = \sqrt{3}$. Therefore,

$$\boxed{\nu = \frac{\log \sqrt{3}}{\log(3/2)} \approx 1.355.} \quad (7.62)$$

The *critical exponents* like ν are constant across all members of a universality class. However, other quantities, like the locations of the critical thresholds, are typically model-specific. This means that ν will be the same on a square lattice, for example, while p_c will be different (in fact, it is roughly 0.593).

In order to illustrate the RG concepts, this section has worked through an example

where the theory is fairly straightforward to apply. It turns out that had we chosen to start with a square lattice instead of a triangular one, our task would have been significantly more difficult, and we'd have to sweat a bit more even to get an *approximation* for ν and for p_c .

We have seen RG theory at work in two rather different contexts. It can be deployed in many more, and I am not aware of a comprehensive reference that covers them all, or even the ones that are by now well-established. Introductory texts include those by Creswick, Farach and Poole [250] and by McComb [251]. The textbooks by Kardar [199] and by Zee [252] cover some aspects of RG in field theory. Bar-Yam [253] provides a succinct first course on RG that includes its classic use in chaos theory. The admirably slim volume by Cardy [254] includes a study of directed percolation, which is relevant to our concerns here.

7.3 Doi Formalism

The next step is to write stochastic dynamical processes in such a way that RG theory is applicable. We can employ the operator tools pioneered by Masao Doi [111, 112, 166, 255, 256, 257, 258] to do so. After some preliminaries, we will be able to use this formalism to describe evolutionary ecological systems.

We can take our system to consist of a series of components indexed by i , where i might range over individuals, genes or sites of a lattice or other network, as appropriate. The state of the i -th component will be denoted x_i ; we can then represent the configuration of the entire system as $\underline{x}$. The probability of being in the state $\underline{x}$ can increase if the system can transition into it, and it can decrease if transitions can take the system out of it. We thus write the master equation for the system dynamics:

$$\frac{dp(\underline{x})}{dt} = \sum_{\underline{y}} \pi(\underline{y} \rightarrow \underline{x}) p(\underline{y}) - \sum_{\underline{y}} \pi(\underline{x} \rightarrow \underline{y}) p(\underline{x}). \quad (7.63)$$

The form of the transition rates $\pi(\underline{x} \rightarrow \underline{y})$ depends upon the detailed interactions of the system—how hosts reproduce, how infections are transmitted and so forth. We can make the master equation Eq. (7.63) tractable if we make a simplifying approximation; of course, any such specialization carries its own cost in biological realism.

The mean-field simplification of Eq. (7.63) is

$$\frac{dp(x)}{dt} = \sum_y \pi(y \rightarrow x) p(y) - \sum_y \pi(x \rightarrow y) p(x). \quad (7.64)$$

In Eq. (7.63), the “vector” $\underline{x}$ labeled a microstate; in Eq. (7.64), x without the arrow labels a macrostate to which many distinct microstates map under a coarse-graining transformation.

To take a concrete example, consider a Lotka–Volterra model of hosts (prey) and consumers (predators). The number of hosts, H , changes through birth and predation, while the number C of consumers increases through predation and decreases due to

7 Spatial Stochastic Mechanics

death. The differential equation for the probability $p(C, H)$ will have a predation term proportional to $p(C - 1, H + 1)$ and to the consumer voraciousness λ , since each consumption act must increase C by 1 and decrement H by the same amount. By the same reasoning, the consumer death term must be proportional to $p(C + 1, H)$ and to the death rate μ ; likewise, prey reproduction generates a term involving $p(C, H - 1)$ and the growth rate σ . Finally, we must have a negative term reflecting how $p(C, H)$ decreases due to death, birth and predation “taking probability away” from the (C, H) macrostate. We can thus write the mean-field master equation [99]

$$\begin{aligned} \frac{dp(C, H)}{dt} = & \lambda(C - 1)(H + 1)p(C - 1, H + 1) \\ & + \mu(C + 1)p(C + 1, H) \\ & + \sigma(H - 1)p(C, H - 1) \\ & - (\mu C + \sigma H + \lambda CH)p(C, H). \end{aligned} \quad (7.65)$$

To rewrite this equation in a form more accessible to the tools we have from our physics education, we introduce creation and annihilation operators. In this case, we require one pair of operators for each species; more generally, the operators will be indexed by some variable i . Annihilation operators a_i and creation operators $a_j^\dagger$ satisfy the commutation relation

$$[a_i, a_j^\dagger] = \delta_{ij}. \quad (7.66)$$

A person excessively steeped in category theory would say that this is just the natural thing to do when dealing with a set of objects whose size can be incremented or decremented [259]. The reason is that we are really considering the *number of ways to perform some manipulation* on a collection of objects: there are n ways to draw a toy out of a box containing n of them, but only one way to drop a new toy in, and so the operations of “removing a toy” and “adding a toy” fail to commute by one unit. Thanks to the product rule, differentiation and multiplication by a variable x satisfy the commutator in Eq. (7.66). This is ultimately why we can discuss probability distributions using generating functions.

We build up states by acting on the vacuum $|0\rangle$ with creation operators. The state so built will be labeled by the occupation numbers, *i.e.*, by how many times we acted with $a_i^\dagger$, for all allowed values of i .

$$|\underline{n}\rangle = \left(\prod_i (a_i^\dagger)^{n_i} \right) |0\rangle. \quad (7.67)$$

The normalization we have chosen implies that the action of a_i is to lower the label of a state and produce a prefactor. In the simplest case, where we have only one type of object,

$$a |n\rangle = a(a^\dagger)^n |0\rangle = n(a^\dagger)^{n-1} |0\rangle = n |n-1\rangle. \quad (7.68)$$

From this, we can conclude

$$a^\dagger a |n\rangle = n |n\rangle. \quad (7.69)$$

We turn occupation probabilities into state vectors by

$$\begin{aligned} |\phi(t)\rangle &= \sum_{\underline{n}} P(\underline{n}, t) \prod_i \left(a_i^\dagger\right)^{n_i} |0\rangle \\ &= \sum_{\underline{n}} P(\underline{n}, t) |\underline{n}\rangle. \end{aligned} \quad (7.70)$$

Of special importance will be the coherent states, defined via

$$|\underline{\eta}\rangle = \exp\left(\sum_i \eta_i a_i^\dagger\right) |0\rangle. \quad (7.71)$$

These will play an important role in the path integrals studied later. One particular coherent state,

$$|1\rangle = e^{\sum_i a_i^\dagger} |0\rangle, \quad (7.72)$$

is an essential part of calculating probabilities. We know that the scheme for turning state vectors into probabilities cannot be the same as that we use in quantum mechanics; physically, we're working with a wholly classical system, while mathematically, a Dirac bracket $\langle\phi(t)|\mathcal{O}|\phi(t)\rangle$ would be bilinear in the occupation probabilities instead of linear.

If we have only one type of particle, then we only need one number to label a coherent state:

$$|\eta\rangle = e^{\eta a^\dagger} |0\rangle = \sum_{n=0}^{\infty} \frac{\eta^n (a^\dagger)^n}{n!} |0\rangle = \sum_{n=0}^{\infty} \frac{\eta^n}{n!} |n\rangle. \quad (7.73)$$

The *Poisson distribution* is a probability distribution over the nonnegative integers, defined by

$$p_n = \frac{\eta^n e^{-\eta}}{n!}. \quad (7.74)$$

The mean of this distribution is equal to η . (In fact, *all* of the cumulants are equal to η .) So, up to normalization, a coherent state is “an over-educated way of talking about a Poisson distribution” [258]. This will be important later, because we will often want to initialize a stochastic dynamical system with Poissonian starting conditions.

The coherent states defined by Eq. (7.71) are eigenstates of the annihilation operators a_i with eigenvalue η_i ; the proof of this goes through just as that for the quantum harmonic oscillator. Consequently, the adjoint coherent states

$$\langle\underline{\eta}| = \langle 0| \exp\left(\sum_i \eta_i a_i\right) \quad (7.75)$$

are *left* eigenstates of the *creation* operators $a_i^\dagger$. In particular, because $\langle 0| e^{\sum_i a_i}$ is a left eigenstate of $a_i^\dagger$ with eigenvalue unity, we can write the expectation value of an

7 Spatial Stochastic Mechanics

operator $\mathcal{O}$ as

$$\langle \mathcal{O}(t) \rangle = \langle 1 | \mathcal{O} | \phi(t) \rangle. \quad (7.76)$$

It is useful to know the inner product of two coherent states:

$$(\phi_1 | \phi_2) = \exp \left(-\frac{1}{2} |\phi_1|^2 - \frac{1}{2} |\phi_2|^2 + \phi_1^* \phi_2 \right). \quad (7.77)$$

Using these tools, we can rewrite the Lotka–Volterra system of Eq. (7.65) in operator language. To do so, we introduce two sets of creation and annihilation operators, which for brevity we denote with c and h , satisfying

$$[c, c^\dagger] = [h, h^\dagger] = 1, \quad [c, h] = [c, h^\dagger] = 0. \quad (7.78)$$

In terms of these operators, Eq. (7.65) takes the form of an imaginary-time Schrödinger equation:

$$\partial_t |\phi(t)\rangle = -\mathcal{H} |\phi(t)\rangle, \quad (7.79)$$

where

$$\begin{aligned} \mathcal{H} = & \lambda c^\dagger c h^\dagger h + \mu c^\dagger c + \sigma h^\dagger h \\ & - \lambda c^\dagger c^\dagger c h^\dagger h - \mu c + \sigma h^\dagger h^\dagger h, \end{aligned} \quad (7.80)$$

which we can reorganize into

$$\mathcal{H} = \lambda(1 - c^\dagger) c^\dagger c h^\dagger h + \mu(c^\dagger - 1)c + \sigma(1 - h^\dagger) h^\dagger h. \quad (7.81)$$

This example illustrates the general way of constructing *stochastic Hamiltonian operators*, which we can summarize as follows [260]:

[F]or each new particle species additional occupation numbers, second-quantized operators, and fields are to be introduced. The details of the reaction are coded into the master equation, though after some practice, it is actually easier to directly start with the Doi time evolution operator, as it is a more efficient representation. The general result is as follows: For a given reaction, two terms appear in the quasi-Hamiltonian (as in the original master equation). The first contribution, which is positive, contains both an annihilation and creation operator for each reactant, normal-ordered. For example, for the $A + A \rightarrow 0$ and $A + A \rightarrow A$ reactions this term reads $\hat{a}^{\dagger 2} \hat{a}^2$, whereas one obtains for the $A + B \rightarrow 0$ reaction $\hat{a}^\dagger \hat{b}^\dagger \hat{a} \hat{b}$. These contributions indicate that the respective second-order processes contain the particle density products a^2 and ab in the corresponding classical rate equations. The second term in the quasi-Hamiltonian, which is negative, entails an annihilation operator for every reactant and a creation operator for every product, normal-ordered. For example, in $A + A \rightarrow 0$ this term would be $\hat{a}^2$, whereas for $A + A \rightarrow A$ it becomes $\hat{a}^\dagger \hat{a}^2$, and for $A + B + C \rightarrow A + B$ it would read $\hat{a}^\dagger \hat{b}^\dagger \hat{a} \hat{b} \hat{c}$. These terms

thus directly reflect the occurring annihilation and creation processes in second-quantized language.

As mentioned, it becomes fairly easy with practice to write the time-evolution operator directly from the description of the reactions. We will do this for some examples in the next section. Before moving on, however, we note one subtlety which is relevant on occasion when reading papers in this field.

Calculating probabilities using Eq. (7.76) introduces a new wrinkle as we move into more advanced computations, because the standard machinery of field theory expects operators to be *normal-ordered*, that is, to have annihilation operators to the right of creation operators in all products thereof. Since the coherent state $(1|$ involves an exponential of annihilation operators a_i , expressions involving $(1|$ and an operator like the Hamiltonian of Eq. (7.81) will not be normal-ordered. To resolve this, we can commute $e^{\sum_i a_i}$ through the Hamiltonian. Because

$$\begin{aligned} e^a a^\dagger &= \sum_{n=0}^{\infty} \frac{a^n}{n!} a^\dagger \\ &= \sum_{n=0}^{\infty} \frac{a^\dagger a^n + n a^{n-1}}{n!} \\ &= (1 + a^\dagger) e^a, \end{aligned}$$

the effect of commuting the exponential through the operator being bracketed is to shift $a_i^\dagger \rightarrow a_i^\dagger + 1$.

7.4 Examples

First, let's apply the procedure described above to the reaction $A \rightarrow 0$, which we shall say happens at rate λ . The Hamiltonian operator is

$$\mathcal{H} = \lambda(a^\dagger a - a) = \lambda(N - a). \quad (7.82)$$

We would like to know how the expected number of particles present, $\langle N \rangle$, will change with time. From Eqs. (7.76) and (7.79), we have

$$\frac{d}{dt} \langle N \rangle = - \langle 1 | N \mathcal{H} | \phi(t) \rangle \quad (7.83)$$

$$= -\lambda \langle 1 | N(N - a) | \phi(t) \rangle. \quad (7.84)$$

At this juncture, it is useful to note a couple relations that follow from the actions of the a and $a^\dagger$ operators on the vectors $|n\rangle$:

$$(1| a^\dagger | \phi \rangle = (1| \phi \rangle, \quad (7.85)$$

$$(1| a | \phi \rangle = (1| N | \phi \rangle = \langle N \rangle. \quad (7.86)$$

7 Spatial Stochastic Mechanics

In addition, the fundamental commutator between a and $a^\dagger$ implies that

$$[a, N] = a. \quad (7.87)$$

Now, we compute:

$$\lambda (1|N(a - N)|\phi) = \lambda (1|(Na - N^2)|\phi) \quad (7.88)$$

$$= \lambda (1|(aN - N - N^2)|\phi) \quad (7.89)$$

$$= \lambda (1|(N^2 - N - N^2)|\phi) \quad (7.90)$$

$$= -\lambda (1|N|\phi). \quad (7.91)$$

Therefore,

$$\boxed{\frac{d}{dt} \langle N \rangle = -\lambda \langle N \rangle.} \quad (7.92)$$

We have shown that the reaction $A \rightarrow 0$ implies *exponential decay*.

Likewise, we can see that the reproduction reaction $A \rightarrow A + A$ produces exponential growth. The stochastic Hamiltonian for this reaction is

$$\mathcal{H} = \lambda [a^\dagger a - (a^\dagger)^2 a]. \quad (7.93)$$

As before, we find the time derivative of the expectation value $\langle N \rangle$:

$$\frac{d}{dt} (1|N|\phi) = -\lambda (1|N [a^\dagger a - (a^\dagger)^2 a]|\phi). \quad (7.94)$$

The expression in the middle is equivalent to

$$N [N - (a^\dagger)^2 a] = N^2 - Na^\dagger N. \quad (7.95)$$

We can commute the $a^\dagger$ through the number operator N , producing

$$N [N - (a^\dagger)^2 a] = N^2 - a^\dagger (N + 1)N. \quad (7.96)$$

Therefore,

$$-\lambda (1|[N^2 - a^\dagger (N + 1)N]|\phi) = \lambda (1|[(N + 1)N - N^2]|\phi) \quad (7.97)$$

$$= \lambda (1|N|\phi). \quad (7.98)$$

And we see that indeed

$$\frac{d}{dt} \langle N \rangle = \lambda \langle N \rangle. \quad (7.99)$$

We conclude this section with an example in which there are multiple locations at which particles can be present. Suppose that we have two boxes, which we call box 1 and box 2, and particles can hop stochastically from one site to the other. We can treat this as a particle being annihilated in box 1 and a replacement being created

within box 2. If this takes place at a rate λ , then

$$\mathcal{H}_{1 \rightarrow 2} = \lambda(a_1^\dagger - a_2^\dagger)a_1. \quad (7.100)$$

For the reverse process,

$$\mathcal{H}_{2 \rightarrow 1} = \lambda(a_2^\dagger - a_1^\dagger)a_2. \quad (7.101)$$

If we let transitions happen in both directions,

$$\mathcal{H} = \mathcal{H}_{1 \rightarrow 2} + \mathcal{H}_{2 \rightarrow 1} = \lambda(a_1^\dagger a_1 - a_2^\dagger a_1 + a_2^\dagger a_2 - a_1^\dagger a_2), \quad (7.102)$$

which we can factor as

$$\mathcal{H} = \lambda(a_1^\dagger - a_2^\dagger)(a_1 - a_2). \quad (7.103)$$

Generalizing this to a large set of boxes, we can write a stochastic Hamiltonian for hopping between adjacent locations, in terms of a sum over nearest-neighbor sites:

$$\mathcal{H} = \lambda \sum_{\langle k j \rangle} (a_k^\dagger - a_j^\dagger)(a_k - a_j). \quad (7.104)$$

This models a *diffusion* process, if we take the rate λ as given by a diffusion coefficient D , divided by the square of an inter-site spacing Δx .

7.5 Coherent-State Path Integrals

Path integrals in field theory are much like partition functions in statistical mechanics. In both cases, we take a sum over terms, each of which is an exponential of a formula encoding the system dynamics, the sum running over all possible ways the system can do something. When we calculated a partition function in elementary statistical mechanics, we summed over all the *states* of a system; when we compute a path integral, we sum over all *paths* which the system can take from one state to another.

Start with some initial state $|\phi(0)\rangle$, which we produce by acting on the vacuum $|0\rangle$ with some combination of raising operators. This state evolves in time according to Eq. (7.79); at some later time t , we compute the expectation value of some operator $\mathcal{O}$:

$$\langle \mathcal{O}(t) \rangle = \langle 1 | \mathcal{O} e^{-\mathcal{H}t} | \phi(0) \rangle. \quad (7.105)$$

We could find the same expectation value by time-advancing the system in M repeated small increments of size $\Delta t = t/M$.

$$\exp(-\mathcal{H}t) = \lim_{M \rightarrow \infty} (1 - \mathcal{H}\Delta t)^M. \quad (7.106)$$

Various ways of developing the path integral all come from different choices for writing the options available to the system at each timestep. This amounts to a choice of basis. We have a long string of factors of $(1 - \mathcal{H}\Delta t)$, each one denoting the advance from t_i to t_{i+1} , where $t_i = i\Delta t$. At time t_i , a system in any one state has some probability

7 Spatial Stochastic Mechanics

of transitioning into any other state; a *path* is a set of M such transitions. The normalization of our basis states gives us a “resolution of unity”: between every two successive factors of $(1 - \mathcal{H}\Delta t)$, we insert a sum over expressions of the form $|x\rangle\langle x|$ which works out to 1. In our case, we employ the coherent-state resolution of unity,

$$\int \frac{d\tilde{\eta}d\eta}{\pi^N} |\underline{\eta}\rangle \langle \underline{\eta}| = 1. \quad (7.107)$$

The factors of π are introduced to cancel those which arise from Gaussian integrals over the state labels. (Recall the normalization of the basic Gaussian curve in Eq. (5.147).)

The *coherent-state representation* of an operator $\mathcal{O}$ is

$$\mathcal{O}[\tilde{\psi}(t), \phi(t)] \equiv \langle \psi(t) | \mathcal{O}(t) | \phi(t) \rangle. \quad (7.108)$$

When we insert the coherent-state resolution of the identity into our formal solution for $\langle \mathcal{O}(t) \rangle$, we end up with an integral of the form

$$\int \cdots |\phi_{t+\Delta t}\rangle \langle \phi_{t+\Delta t}| e^{-\mathcal{H}\Delta t} |\phi_t\rangle \langle \phi_t| e^{-\mathcal{H}\Delta t} |\phi_{t-\Delta t}\rangle \langle \phi_{t-\Delta t}| \cdots, \quad (7.109)$$

where we are integrating over $t/\Delta t$ different variables. Each factor in the long integrand is of the form

$$\langle \phi_t | e^{-\mathcal{H}\Delta t} | \phi_{t-\Delta t} \rangle = e^{-\mathcal{H}(\phi_t^*, \phi_{t-\Delta t})\Delta t} \langle \phi_t | \phi_{t-\Delta t} \rangle. \quad (7.110)$$

Using the overlap between coherent states, Eq. (7.77), this becomes

$$\langle \phi_t | e^{-\mathcal{H}\Delta t} | \phi_{t-\Delta t} \rangle = e^{-\mathcal{H}(\phi_t^*, \phi_{t-\Delta t})\Delta t} \exp\left(-\frac{1}{2} |\phi_t|^2 - \frac{1}{2} |\phi_{t-\Delta t}|^2 + \phi_t^* \phi_{t-\Delta t}\right). \quad (7.111)$$

We can approximate this by

$$\langle \phi_t | e^{-\mathcal{H}\Delta t} | \phi_{t-\Delta t} \rangle \approx e^{-\mathcal{H}(\phi_t^*, \phi_t)\Delta t} \exp\left(-\frac{1}{2} |\phi_t|^2 - \frac{1}{2} |\phi_{t-\Delta t}|^2 + \phi_t^* \phi_{t-\Delta t}\right). \quad (7.112)$$

And, by considering the terms produced by adjacent factors in the big integrand, we can simplify this to

$$\langle \phi_t | e^{-\mathcal{H}\Delta t} | \phi_{t-\Delta t} \rangle \approx e^{-\mathcal{H}(\phi_t^*, \phi_t)\Delta t} \exp(-\phi_t^* \partial_t \phi_t \Delta t). \quad (7.113)$$

Piling up many time slices and taking the limit $\Delta t \rightarrow 0$, we get that e^{-Ht} becomes

$$\int \mathcal{D}\phi^* \mathcal{D}\phi \exp\left(-\int_0^t dt' [\phi^* \partial_{t'} \phi + H(\phi^*, \phi)]\right). \quad (7.114)$$

In this integral, the product of the integration volumes $d^2\phi_j/\pi$ has become $\mathcal{D}\phi^* \mathcal{D}\phi$.

The *action* which determines the weighting of each trajectory in the path integral

is

$$S[\phi^*, \phi] = \int_0^t dt' [\phi^* \partial_{t'} \phi + H(\phi^*, \phi)]. \quad (7.115)$$

We can include the initial condition $|\phi(0)\rangle$ and the projection state $\langle 1|$ by adding a couple terms to the action. For example, if we start with a Poisson distribution with mean n_0 , then

$$S[\phi^*, \phi] = \int_0^t dt' [\phi^* \partial_{t'} \phi + H(\phi^*, \phi)] - \phi(t) - n_0 \phi^*(0). \quad (7.116)$$

Now, the expectation value $\langle \mathcal{O}(t) \rangle$ is

$$\langle \mathcal{O}(t) \rangle = \frac{1}{\mathcal{N}} \int D\phi_j^* D\phi_j \mathcal{O}(\phi(t)) e^{-S[\phi^*, \phi]}. \quad (7.117)$$

The prefactor is a normalization constant that won't concern us.

We can eliminate the term that was brought in by the projection state $\langle 1|$, by defining a new variable:

$$\phi^* \rightarrow 1 + \tilde{\phi}. \quad (7.118)$$

The time-derivative term inside the action integral becomes

$$\int_0^t dt' (1 + \tilde{\phi}) \partial_{t'} \phi = \phi(t) - \phi(0) + \int_0^t dt' \tilde{\phi} \partial_{t'} \phi. \quad (7.119)$$

This shift also transforms H , taking it to $H(1 + \tilde{\phi}, \phi)$.

If we have multiple lattice sites, then the action is a function of all the ϕ_k and the $\tilde{\phi}_k$:

$$S[\{\tilde{\phi}_k\}, \{\phi_k\}] = \sum_k \left[\int_0^t dt' [\tilde{\phi}_k \partial_{t'} \phi_k + H(\{\tilde{\phi}_k\}, \{\phi_k\})] - n_0 \tilde{\phi}_k(0) \right]. \quad (7.120)$$

Our notation in this section largely follows the lectures by Vollmayr-Lee [261], which provide some additional details.

7.6 Spatial Dependence

Incorporating spatial extent into a model means promoting our creation and annihilation operators to sets thereof, indexed by spatial position, and adding appropriate movement terms to the stochastic Hamiltonian. This procedure creates an expression that can be studied by means of RG theory.

Most of the work done on this topic assumes that particles move about by diffusion and react in some way if they happen to bump into one another. Therefore, we start with a mathematical model of a diffusion process. The diffusion action, assuming Poisson initial conditions, is just the diffusion Hamiltonian (7.104) in our new coherent-

7 Spatial Stochastic Mechanics

state language:

$$S_D = \int dt \left[\sum_i \tilde{\phi}_i \partial_{t'} \phi_i + \frac{D}{\Delta x^2} \sum_{\langle ij \rangle} (\tilde{\phi}_i - \tilde{\phi}_j)(\phi_i - \phi_j) \right] - \sum_i n_0 \tilde{\phi}_i(0). \quad (7.121)$$

If we take the continuum limit, ϕ and $\tilde{\phi}$ become fields, and the action becomes

$$S_D = \int dt d^d x \left[\tilde{\phi} \partial_{t'} \phi + D \nabla \tilde{\phi} \cdot \nabla \phi - n_0 \tilde{\phi} \delta(t) \right]. \quad (7.122)$$

Integrating by parts, we find

$$S_D = \int dt d^d x \left[\tilde{\phi} (\partial_{t'} - D \nabla^2) \phi - n_0 \tilde{\phi} \delta(t) \right]. \quad (7.123)$$

This action is linear in the response field $\tilde{\phi}$. Therefore, finding its extremum is easy:

$$\frac{\delta S_D}{\delta \tilde{\phi}} = \partial_t \phi - D \nabla^2 \phi - n_0 \delta(t) = 0. \quad (7.124)$$

This is satisfied when

$$\partial_t \phi = D \nabla^2 \phi + n_0 \delta(t), \quad (7.125)$$

which is the familiar diffusion equation, with initial conditions specified at time $t = 0$.

Mobilia *et al.* look in particular at the sustainability transition, where the carnivores are consuming just enough to keep up their numbers and not go extinct [99, 262]. After a lot of redefinitions and rescalings, they get an action which is an integral of a Lagrangian density of the form

$$\mathcal{L}[\tilde{\psi}, \psi] = \tilde{\psi} (\partial_t + D_c(r_c - \nabla^2)) \psi - u \tilde{\psi} (\tilde{\psi} - \psi) \psi + \tau \tilde{\psi}^2 \psi^2. \quad (7.126)$$

The field ψ is a shifted and rescaled carnivore density, and $\tilde{\psi}$ is its conjugate “response field”. The various constants are mixed-together combinations of the parameters we started with.¹ There are a fair many assumptions in the derivation of this Lagrangian which ought to be unpacked (such as the decision to expand perturbatively around diffusive motion), and of course, once we’ve squeezed the dynamics we want to study into this formalism, we’d like to get answers *out* of it, and *that* appears to involve a whole lot of elaborate RG machinery.

¹This is, incidentally, the same form of Lagrangian which defines the “Reggeon field theory” used back in the ’70s for high-energy scattering physics, with the last term providing what I think they’d call a quadruple-Pomeron vertex [263, 264, 265].

7.7 Rate Equations from Tree-Level Calculations

In this field-theoretic formalism, rate equations arise as tree-level results, *i.e.*, the results obtained by using Feynman diagrams containing no closed loops. This is easiest to appreciate in a simple model of diffusion-limited annihilation. In this model, particles move about by Brownian motion, and if they come into contact, they annihilate with some probability. Feynman diagrams for this model have a natural interpretation in terms of ways a particle can survive over an interval of time.

Suppose we release a particle at time 0 and find it again at some later time t . What could have happened to the particle in between, given that it survived until t ? The simplest possibility is that it encountered no other particles and so had no opportunity to annihilate. The diagram for this process is just a straight line. Alternatively, the particle could have bumped into one or more other particles but failed to annihilate each time. Diagrams for these processes have vertices where two lines come in and one or two lines come out. The total probability of surviving until time t is the sum over all the ways the particle could survive, which means a sum over all possible diagrams.

Neglecting correlations means neglecting the possibility that two particles coming together at a certain time have done so already in the past. This means that neglecting correlations allows us to omit all diagrams containing closed loops from our sum. This calculation is significantly easier than the complete version. Figure 7.3 shows the reason in diagrammatic form: the heavy line, representing all possible tree-shaped diagrams, obeys a self-consistency condition.

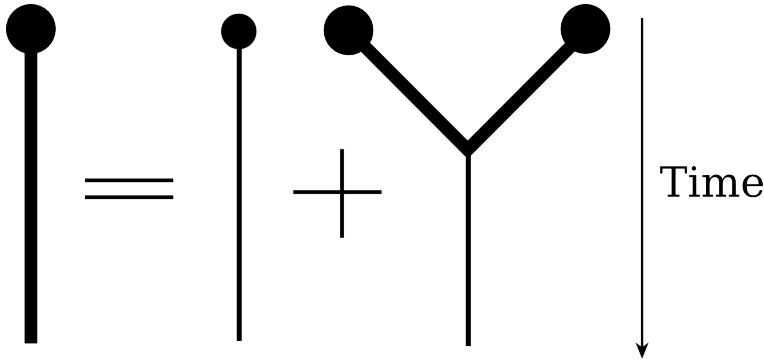


Figure 7.3: Self-consistency calculation for the diffusion-limited annihilation model's tree-level propagator.

The *propagator* tells us how particles get from one point to another if nothing happens in between. We're saying that these particles move by diffusion, so the propagator in this case is the Green's function for the diffusion equation.

$$G_D(k, \omega) = \frac{1}{-i\omega + Dk^2}. \quad (7.127)$$

7 Spatial Stochastic Mechanics

Transforming from frequency back into the time domain,

$$G_D(k, t) = \int \frac{d\omega}{2\pi} \frac{e^{-i\omega t}}{-i\omega + Dk^2}. \quad (7.128)$$

For $t > 0$, this is

$$G_D(k, t) = \exp(-Dk^2 t). \quad (7.129)$$

In position space,

$$G_D(x, t > 0) = \frac{e^{-x^2/(4Dt)}}{(4\pi Dt)^{d/2}}. \quad (7.130)$$

We can think of this as saying that the response to a delta-function spike at $t = 0$ is a Gaussian curve which spreads out as time passes, its standard deviation growing as the square root of the elapsed time.

To each trivalent vertex, we associate a factor $-2\lambda_0$, and each initial vertex gets a n_0 . Wave-vector (or “momentum”) conservation applies at each vertex. We can read off the self-consistency condition for the tree-level contributions directly from the diagrams:

$$a_{tree}(t) = n_0 + \int_0^t dt_1 G_D(0, t - t_1) (-2\lambda_0) a_{tree}(t_1)^2. \quad (7.131)$$

The propagator with $k = 0$ is just 1. Differentiating both sides of the self-consistency equation yields that the time derivative of a_{tree} is the integrand evaluated at t .

$$\frac{da_{tree}}{dt} = -2\lambda_0 a_{tree}^2 \quad (7.132)$$

This is just a rate equation for a_{tree} . With the initial condition $a_{tree}(0) = n_0$, this has the solution

$$a_{tree}(t) = \frac{n_0}{1 + 2\lambda_0 n_0 t}. \quad (7.133)$$

Going beyond tree-level calculations requires RG theory. The computations are, at least for diffusion-limited annihilation, not all that difficult on a technical level, once the basic concepts are grasped. One fact which enables considerable simplifications is that the reactions in this model cannot increase the number of particles present. Applying RG to models in which this condition does not hold is more complicated.

Question: *Since we require the concepts of RG at such an early stage of our analysis, would a more sophisticated mathematical treatment of that theory be useful? It appears that the Hopf-algebraic study of RG [266] would be just as applicable in this stochastic setting as it is in QFT. Perhaps diffusion-limited annihilation might even provide a simpler setting for it.*

7.8 Directed Percolation

Back in Chapter 3, we introduced directed percolation as an idealized model of fluid flow through a porous medium. We began with the mental image of a regular lattice of channels, some of whose junction points were blocked. If the fraction of blocked points was too large, fluid flowing through the lattice from top to bottom would always be stymied, but if the blockages were sparse, the fluid could percolate downwards through the channels. We can treat this model in any number of dimensions; the crucial point is that there is a preferred direction. And because the fluid only spreads downhill, there is another picture available. We can also think of the tip of each stream as a corpuscle executing a *random walk*. These walkers can merge, if two fluid streams come together at a common point, and they can reproduce, which happens when a fluid stream splits and its offshoots progress along two different channels. The walkers can also perish: this corresponds to a stream which encounters a point where no further propagation is possible.

In short, directed percolation in a $(d + 1)$ -dimensional lattice, with one dimension singled out as the axis of gravity, can equally well be thought of as *merging, reproducing and perishing random walks* in a d -dimensional space. This latter conception of the problem is amenable to the techniques we have developed in this chapter.

Consider the spatial stochastic process defined by the combination of particle decay ($A \rightarrow 0$), reproduction ($A \rightarrow A + A$) and merging ($A + A \rightarrow A$) with diffusive motion. Call the decay or mortality rate μ , the reproduction rate σ and the merging rate λ .

The stochastic action is

$$S = \int d^d x \left[-\phi(t) + \int_0^t dt' \left(\phi^* (\partial_{t'} - D \nabla^2) \phi - \mu(1 - \phi^*) \phi + \sigma(1 - \phi^*) \phi^* \phi - \lambda(1 - \phi^*) \phi^* \phi^2 \right) - n_0 \phi^*(0) \right]. \quad (7.134)$$

Performing the field shift $\phi^* = 1 + \tilde{\phi}$, and writing $r = (\mu - \sigma)/D$, we obtain

$$S = \int d^d x dt' \left[\tilde{\phi} (\partial_{t'} + D(r - \nabla^2)) \phi - \sigma \tilde{\phi}^2 \phi + \lambda \tilde{\phi} \phi^2 + \lambda \tilde{\phi}^2 \phi^2 \right]. \quad (7.135)$$

We have deliberately dropped the initial-conditions term, because particle decay and generation will scramble the original configuration. Note that $\delta S / \delta \phi = 0$ is satisfied by $\tilde{\phi} = 0$, and combining this with the other variation $\delta S / \delta \tilde{\phi} = 0$ yields the equation of motion

$$\partial_t \phi = -D(r - \nabla^2) \phi - \lambda \phi^2. \quad (7.136)$$

If particles can *annihilate* as well as merge ($A + A \rightarrow 0$), then the action is modified:

$$S = \int d^d x dt' \left[\tilde{\phi} (\partial_{t'} + D(r - \nabla^2)) \phi - \sigma \tilde{\phi}^2 \phi + (\lambda + 2\lambda') \tilde{\phi} \phi^2 + (\lambda + \lambda') \tilde{\phi}^2 \phi^2 \right]. \quad (7.137)$$

And the rate equation becomes

$$\partial_t \phi = -D(r - \nabla^2)\phi - (\lambda + 2\lambda')\phi^2. \quad (7.138)$$

It is interesting to follow the history of the directed-percolation concept. It was first proposed in 1957 [267]. The mathematics necessary to treat it cleverly was invented (or, rather, adapted from a different area of physics) in the 1970s, and then forgotten, and then rediscovered by somebody else [257]. Connections with other subjects were made. Experiments were carried out on systems which *almost* behaved like the idealization, but always turned out to differ in some way... until 2007, when the behavior was finally caught in the wild [268]. And this experiment, which at last observed a DP-class phase transition with quantitative exactness, used a liquid crystal substance (*N*-(4-Methoxybenzylidene)-4-butylaniline) which wasn't synthesized until 1969 [269].

This rather puts the lie to the notion that a scientific hypothesis must be amenable to immediate falsification by experiment. The model here is not an esoteric proposal for quantum gravity, but an idealization of water flowing through coffee grounds. And not only did it take half a century to go from a mathematician's mind to a laboratory bench, but that journey depended on tools which did not exist when the model was first conceived.

7.9 Prior Relevant Results and Difficulties

One can find in the literature various applications of the Doi field-theoretic formalism to systems which resemble our host-consumer model. Field-theory tools allow one to calculate, for example, the critical exponents which describe the dynamics of a predator population near its extinction threshold [99, 260]. They also appear to be fairly successful at locating where the critical point will occur in an epidemic model, in at least some regions of parameter space [270].

One significant problem blocks the path to applying these techniques to the spatial host-consumer model studied in this report. All the prior work assumed *diffusive motion*, as we did in our brief encounter with annihilating particles above. The host-consumer model has no such feature: all motion is due to *reproduction* into adjacent lattice sites (empty sites for hosts, host-occupied sites for consumers). The validity of expanding around a diffusion propagator is, therefore, questionable. It may be a viable approximation in some cases, however, thanks to universality. In a later section, we will make an argument to this effect in more depth.

Simulations such as those depicted in Figure 3.1 suggest that treating the advances of host and consumer populations as *surface growth* may be a useful approach. The field-theoretic study of the Kardar-Parisi-Zhang model may, therefore, be an area to draw from [199, 252]. (Visually, a consumer wave eating its way through a field of hosts does resemble burning paper, which has been studied as a possible instance of KPZ-class dynamics.)

7.10 Carrying Capacity

Realistic ecosystem models typically have some notion of a *carrying capacity*: a given environment can only sustain a living population of limited size. In our spatial host–consumer model, population density is limited by the fact that each lattice site can only hold one H-type or one C-type individual. (As we discussed in Chapter 3, each H-type or C-type agent in the model can be thought of as an idealization of a homogeneous subpopulation.) To make progress, we need to incorporate carrying capacity into the field-theoretic formalism developed above. One way to do this is to introduce rate-limiting terms into the Lagrangian, which ensure that the probability of the population density growing too large at any point in space is negligible. Another way is to curtail the state space of the theory itself. This latter approach relates to some interesting topics in mathematical physics, so we’ll explore it at greater length.

In the Schwinger oscillator model, we have two sets of harmonic oscillator creation and annihilation operators,

$$[a_-, a_-^\dagger] = 1, \quad [a_+, a_+^\dagger] = 1, \quad (7.139)$$

such that operators pertaining to one oscillator commute with those for the other:

$$[a_+, a_-^\dagger] = [a_-, a_+^\dagger] = 0. \quad (7.140)$$

Excited states—that is, states with nonzero particle number—are built by acting repeatedly with the creation operators:

$$|n_+ n_-\rangle = \frac{(a_+^\dagger)^{n_+} (a_-^\dagger)^{n_-}}{\sqrt{n_+!} \sqrt{n_-!}} |0, 0\rangle. \quad (7.141)$$

The number operators $N_+ = a_+^\dagger a_+$ and $N_- = a_-^\dagger a_-$ are eigenvalues of $|n_+ n_-\rangle$ with eigenvalues n_+ and n_- , respectively. Under the substitution

$$n_+ \rightarrow j + m, \quad n_- \rightarrow j - m, \quad (7.142)$$

we can write

$$|j, m\rangle = \frac{(a_+^\dagger)^{j+m} (a_-^\dagger)^{j-m}}{\sqrt{(j+m)!} \sqrt{(j-m)!}} |0, 0\rangle. \quad (7.143)$$

Defining

$$J_+ = \hbar a_+^\dagger a_-, \quad J_- = \hbar a_-^\dagger a_+, \quad J_z = \frac{\hbar}{2} (N_+ - N_-), \quad (7.144)$$

the operators $J_\pm$ and J_z satisfy the angular momentum algebra $\mathfrak{su}(2)$:

$$[J_z, J_\pm] = \pm J_\pm, \quad (7.145)$$

$$[J_+, J_-] = 2J_z. \quad (7.146)$$

To relate Schwinger’s work and stochastic mechanics, take a system composed of

7 Spatial Stochastic Mechanics

two particle species, which we can call type a and type b . The different normalization does not affect the operator algebra. We can still define $J_+ = a^\dagger b$, $J_- = b^\dagger a$ and $J_z = (a^\dagger a - b^\dagger b)/2$. Acting on a ket labeled by particle number, we still have

$$J_z |n_a, n_b\rangle = \frac{1}{2}(n_a - n_b) |n_a, n_b\rangle, \quad (7.147)$$

and the commutators among our operators are again

$$[J_z, J_\pm] = \pm J_\pm, \quad (7.148)$$

$$[J_+, J_-] = 2J_z. \quad (7.149)$$

Defining

$$J^2 = J_z^2 + \frac{1}{2}(J_+ J_- + J_- J_+), \quad (7.150)$$

we find that

$$J^2 |n_a, n_b\rangle = \frac{N}{2} \left(\frac{N}{2} + 1 \right) |n_a, n_b\rangle, \quad \text{where} \quad N = n_a + n_b. \quad (7.151)$$

This means that we can label basis kets just as well by j and m as we could by n_a and n_b .

If the total number of individuals present is N , then we are working with the N -dimensional representation of $\mathfrak{su}(2)$.

In the host–consumer model, each lattice site can be in one of three states. If we write H_i , C_i and E_i for the number of hosts, consumers and empty slots at position i , then

$$H_i + C_i + E_i = 1. \quad (7.152)$$

Introducing the number operator

$$N_i = h_i^\dagger h_i + c_i^\dagger c_i + e_i^\dagger e_i, \quad (7.153)$$

we have that N_i takes the value 1 on all admissible states. N_i commutes with all nine operators which annihilate an individual and then create one: $h_i^\dagger h_i$, $h_i^\dagger c_i$, $c_i^\dagger h_i$ and so on. So, N_i commutes with all linear combinations of operators of this form. We have, then, an eight-dimensional space of operators, which is the Lie algebra $\mathfrak{su}(3)$, in Schwinger form.²

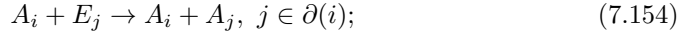
7.11 A Common Denominator

One stochastic model recurs in many of the settings we have explored in different chapters of this thesis. This model is a stochastic process which includes *reproduction by budding into empty space* and *death leaving empty space behind*, implemented on

²Actually, it's $\mathfrak{sl}(3)$, the complexification of $\mathfrak{su}(3)$. See [271].

a lattice or other network. By understanding this stochastic process, we can obtain quantitative predictions for the systems we studied in Chapters 3 and 4.

We can identify the basic reactions in this model as follows. Let A_i denote the presence of a particle at site i , and let E_i denote the fact of site i being empty. Writing i and j for site indices, and $\partial(i)$ for the neighborhood of site i , we have



The first reaction is reproduction; and the second, death. Each of these processes happens with some probability whenever the conditions are right to allow it, *i.e.*, whenever the appropriate reactants are present in the proper juxtaposition. Together, these reactions define a kind of *contact process*.

The eco-evolutionary systems that have motivated us to write this simplified model typically include higher-order complications. The rates at which organisms can die and empty sites can be filled will, in general, be affected by the surroundings. Therefore, we include one generalization beyond the standard contact process: The probability of the reaction taking place, when it is allowed, can depend on the local environment. Specifically, we are concerned with cases where the probability of a particle at site i reproducing to make a new particle at site j depends on all the sites which are adjacent to i . This means that the probability of site j becoming occupied can depend on the *next-to-nearest* neighbors of j , which is unlike the basic contact process.

This model includes the single-species cases of the Volunteer's Dilemma, which we studied in §4.2. When the lattice contains only Volunteers or only Slackers, the only difference between the two species' basic dynamical rules are how the presence of neighboring individuals affect the rate of the budding process. Furthermore, this model can serve to approximate the spatial host–consumer model of Chapter 3, in a scenario where the ecosystem is filled with hosts, and only a small number of consumers are present in a localized area. When the consumer transmissibility is just barely large enough to keep the consumer population from fizzling, and the lattice contains no empty space, we can gain some understanding from a simplified model which has only two possible states for each lattice site. In essence, we can to a first approximation neglect the possibility of empty sites, as long as the host growth rate is nonzero, because any sites that are emptied will be refilled quickly enough not to be of concern.

The empty lattice is an absorbing state for this model: fluctuations can take us into it, but never out. Experience in nonequilibrium statistical mechanics, codified in a conjecture of Grassberger and Janssen [272, 273], suggests that the phase transition between the active and absorbing regimes will belong to the directed percolation universality class. Indeed, if our system had diffusive particle motion and no occupation restrictions, we could make that identification immediately. Likewise, the transition in the standard contact process is known to be a DP-class phenomenon [121].

The question now arises: does the modification we made to the contact process affect the universality class to which the model belongs? If we can argue that it does *not*, then we can use off-the-shelf results about DP-class transitions for the spatial host–consumer model and for the Volunteer's Dilemma. This would represent a substantial

7 Spatial Stochastic Mechanics

advance: even if the computation of DP critical exponents is a laborious calculation, it only has to be done for one system.

We now develop two ways of writing equations for stochastic contact processes. This will help tie the model we have defined here into the prior literature. First, we will apply as directly as possible the tools of raising and lowering operators we developed in earlier sections.

If we were not concerned about the need for empty space, we could write the following stochastic Hamiltonians to encapsulate the reactions we described above:

$$H_d = \sum_i \lambda_i (a_i^\dagger a_i - a_i), \quad (7.156)$$

$$H_r = \sum_{\langle i,j \rangle} \sigma_i (a_i^\dagger a_i - a_i^\dagger a_j^\dagger a_i). \quad (7.157)$$

Here, we have introduced the potentially site-dependent transition probabilities σ_i and λ_i . In the simplest case, these can be taken as constant across all sites, but more generally, they will depend on the configuration of other sites in the vicinity of i . That is, whatever higher-order interactions among organisms might arise, we can roll them into the σ_i and λ_i .

However, because each site can hold at most one particle at a time, events which violate that constraint must be disallowed. We can represent this formally by introducing delta functions which make a contribution to the Hamiltonian vanish if conditions are not right. For example, the event of a particle at site i budding to produce an offspring at site j can only take place if exactly one particle exists at i and exactly zero exist at j . So, we write the modified stochastic Hamiltonians,

$$H_r = \sum_{\langle i,j \rangle} \sigma_i (a_i^\dagger a_i - a_i^\dagger a_j^\dagger a_i) \delta_{n_i,1} \delta_{n_j,0}, \quad (7.158)$$

$$H_d = \sum_i \lambda_i (a_i^\dagger a_i - a_i) \delta_{n_i,1}. \quad (7.159)$$

As before, both σ_i and λ_i can be functions of the local environment around site i .

When we pass to the coherent-state representation and field theory, the delta functions in our stochastic Hamiltonians will become exponential factors [274, 275]. This is a consequence of the Fourier representation of the Kronecker delta:

$$\delta_{n,m} = \frac{1}{2\pi} \int dx e^{ix(n-k)}. \quad (7.160)$$

Now for the second approach, in which we incorporate the site-occupation restrictions directly into our definitions of the creation and annihilation operators. Grassberger and de la Torre [276] define the following *simplified Gribov process*, which is known to have a DP-class phase transition. Construct a regular lattice, whose points will be labeled by the index i . Each lattice point can be occupied by at most one particle, *i.e.*, we restrict the occupation number ν_i to be 0 or 1. Each particle can

spontaneously decay with rate κ , and each particle can produce another in an adjacent empty lattice site with rate κ' .

The creation operator for site n acts on a basis state by the rule

$$c_i^\dagger |\cdots \nu_i \cdots\rangle = (1 - \nu_i) |\cdots \nu_i + 1 \cdots\rangle, \quad (7.161)$$

and the annihilation operator on site n acts as

$$c_i |\cdots \nu_i \cdots\rangle = \nu_i |\cdots \nu_i - 1 \cdots\rangle. \quad (7.162)$$

Note that acting twice with either operator destroys the vector. If we apply the combination $c_i^\dagger c_i$ to a state,

$$c_i^\dagger c_i |\cdots \nu_i \cdots\rangle = \nu_i (2 - \nu_i) |\cdots \nu_i \cdots\rangle, \quad (7.163)$$

we see that the result vanishes if $\nu_i = 0$, which is consistent. This justifies using $c_i^\dagger c_i$ as the number operator for site i . Likewise,

$$c_i c_i^\dagger |\cdots \nu_i \cdots\rangle = \nu_i (1 - \nu_i) (1 + \nu_i) |\cdots \nu_i \cdots\rangle, \quad (7.164)$$

which vanishes if $\nu_i = 1$.

Define $|0\rangle$ to be the state annihilated by all c_i , and let

$$V = \prod_i (1 + c_i). \quad (7.165)$$

Then the probability to find particles at locations $n_1, \dots, n_k$ irrespective of what is happening elsewhere on the lattice is

$$\rho^k(n_1, \dots, n_k | \Phi) = \langle 0 | c_{n_1} \cdots c_{n_k} | \Phi \rangle. \quad (7.166)$$

We can check this for a system composed of a single site. Let the state $|\Phi\rangle$ be

$$|\Phi\rangle = (1 - p) |0\rangle + p |1\rangle. \quad (7.167)$$

Then we have that

$$\begin{aligned} \langle 0 | c_1 (1 - c_1) | \Phi \rangle &= \langle 0 | c_1 (1 - c_1) | \Phi \rangle \\ &= \langle 0 | c_1 (1 - p) | 0 \rangle + \langle 0 | c_1 p | 1 \rangle - \langle 0 | c_1^2 (1 - p) | 0 \rangle = \langle 0 | c_1^2 p | 1 \rangle \\ &= p \langle 0 | 0 \rangle \\ &= p. \end{aligned} \quad (7.168)$$

And we see indeed that the probability of finding a particle in the single site of our system is p , as it should be.

7 Spatial Stochastic Mechanics

For a properly normalized state $|\Phi\rangle$, we have

$$\langle 0|V|\Phi\rangle = 1. \quad (7.169)$$

The time evolution of $|\Phi\rangle$ is given by

$$\frac{d}{dt}|\Phi\rangle = -L|\Phi\rangle, \quad (7.170)$$

with the operator L defined as

$$L = \kappa \sum_i (c_i^\dagger - 1) c_i + \frac{\kappa'}{z} \sum_{\langle i,j \rangle} (c_i - 1) c_i^\dagger c_j^\dagger c_j. \quad (7.171)$$

The first sum is over all lattice sites i , and the second is over all pairs of nearest neighbors i and j . In the prefactor of the second sum, we have used z to denote the coordination number of the lattice.

Returning to the first way we set up the problem, let us suppose now that the transition probabilities σ_i and λ_i are independent of position. In the absence of site-occupation restrictions, the stochastic Hamiltonian is

$$H = \lambda \sum_i (a_i^\dagger a_i - a_i) + \sigma \sum_{\langle i,j \rangle} (a_i^\dagger a_i - a_i^\dagger a_j^\dagger a_i) \quad (7.172)$$

Incorporating the restriction that a lattice site can contain at most one particle, the stochastic Hamiltonian becomes

$$H = \lambda \sum_i (a_i^\dagger a_i - a_i) \delta_{n_i,1} + \sigma \sum_{\langle i,j \rangle} (a_i^\dagger a_i - a_i^\dagger a_j^\dagger a_i) \delta_{n_i,1} \delta_{n_j,0}. \quad (7.173)$$

Note that we have defined the basic interactions of our model in terms of pairs; higher-order complications would emerge should we try developing the model in perturbation theory to make numerical computations.

Is there a relationship between L and H ? The annihilation operators a_i are analogous to the c_i , and likewise for $a_i^\dagger$ and $c_i^\dagger$. We saw above that the operator $c_i^\dagger c_i$ acts like a Kronecker delta function, comparing the occupancy of site i to 1. Consequently, $(1 - c_i^\dagger c_i)$ can be thought of as a delta function that compares the population size at site i to 0.

If we replace $\delta_{n_i,1}$ with $c_i^\dagger c_i$, $\delta_{n_i,0}$ with $(1 - c_i^\dagger c_i)$, a_i with c_i and $a_i^\dagger$ with $c_i^\dagger$, then we turn our operator H into the operator we constructed earlier, L . That is, we have established that Eq. (7.171) and Eq. (7.173) are equivalent ways of representing the dynamics of the contact process.

What happens if σ and λ are no longer constant across the system? Does this bump the model out of the DP universality class? We can argue that in fact it does not, for the following reason. Although for brevity we wrote the transition rates as depending on the site index i , we defined them as depending upon the configuration

of particles near i . The value of σ_i can change as the population in the vicinity of site i fluctuates. We could instead have defined a model in which σ_i is chosen at random for each i following some probability distribution, and then all the σ_i remain constant over time. This latter approach is known in the language of statistical physics as *quenched disorder*. And it is quenched disorder that pushes models out of the DP universality class [277, 278, 279]. In contrast, because the fluctuations due to higher-order interactions among organisms are not frozen, they are irrelevant, in the RG sense of the term.

Therefore, we expect that the DP critical exponents will be applicable to the phase transitions in our spatial host–consumer model and the lattice implementation of the Volunteer’s Dilemma. Referring to Figures 3.6, 4.4 and 4.5, it is satisfying to report that the simulations agree.

8 Invasion Fitness by Moment Closure Approximations

8.1 Introduction and Overview

Moment closures are a way of forgetting information about a system in a controlled fashion, in the hope that an incomplete, fairly heavily “coarse-grained” picture of the system will still be useful in figuring out what will happen to it. Sometimes, this is a justifiable hope, but in other cases, we are right to wonder whether all the algebra it generates actually leads us to any insights. Here, we’ll be concerned with a particular application of this technology: studying the vulnerability of an ecosystem to invasion. We shall find expressions for **invasion fitness**, the expected relative growth rate of an initially-rare species or variety.

Consider a lattice, each site of which can be occupied by an individual of “resident” type (R), occupied by a mutant (M), or empty (0). The difference between the mutant-type and resident-type individuals is encoded in the choice of transition rules representing death, birth and migration. We can get an aggregate measure of the situation by finding the probability that a randomly chosen site will be in state a , where a can take values in the set $\{R, M, 0\}$. A finer degree of distinction is provided by the conditional probabilities $q_{a|b}$, where, for example, $q_{R|M}$ denotes the probability that a randomly chosen neighbor site to a randomly chosen mutant is of resident type. Note that if a mutant is injected into a native resident population and its offspring form a geographical cluster, $q_{M|M}$ can be much larger than p_M : few individuals are mutants overall, but the probability of a mutant life-form interacting with another mutant is high.

The *pair dynamics* of the system involves the time evolution of the probabilities p_{ab} , that is, the probability that a randomly selected lattice edge will have a on one end and b on the other. The differential equation for dp_{RM}/dt , for example, will have terms reflecting the processes which can form and destroy RM pairs: $RM \rightarrow RR$ is one possibility, and $RM \rightarrow MM$ is another. Death, which comes for organisms and leaves empty spaces behind, introduces processes like $RM \rightarrow R0$, $RM \rightarrow 0M$ and $RM \rightarrow 00$. Reproduction can lead to formerly empty spaces becoming occupied: $R0 \rightarrow RR$ and $M0 \rightarrow MM$. We’ve moved beyond just creating and annihilating residents and mutants, and now we’re dynamically changing the number of “resident–resident” and “resident–mutant” *pairs*.

Each term in our differential equations will have a transition rate dependent upon a conditional probability of the form $q_{a|bc}$, denoting the probability that a b of a bc pair will have a neighbor of type a . The differential equations for the pair probabilities p_{ab}

8 Invasion Fitness by Moment Closure Approximations

thus depend on triplet probabilities p_{abc} , which depend upon quadruplet probabilities and so forth. To make progress, we truncate this hierarchy, brutally cutting off higher-order correlations by declaring that

$$q_{a|bc} \approx q_{a|b}. \quad (8.1)$$

This imposition, a *pair approximation*, destroys information about spatial structure and thereby introduces bias which in an ideal world ought to be accounted for. In theoretical ecology, this maneuver dates back at least to Matsuda *et al.* in 1992 [109], though it has antecedents in statistical physics, going back to the kinetic theory work of Bogoliubov, Born, Green, Kirkwood and Yvon, for whom the “BBGKY hierarchy” is named.

Invasion fitness is judged in the following manner. We start with a lattice devoid of mutants ($p_{Ma} = 0$) and find the equilibrium densities p_{RR}^* and p_{R0}^* by setting

$$\frac{dp_{R0}}{dt} = \frac{dp_{RR}}{dt} = 0. \quad (8.2)$$

The exact form of p_{RR}^* and p_{R0}^* will depend upon interaction details which we won’t worry about just yet. We then inject a mutant strain into this situation; as the mutants are initially rare, we can say they do not affect the large-scale dynamics of the resident population. Summarizing the pair probabilities p_{Ma} with the shorthand $\underline{p}$, we write the differential equation in matrix form

$$\frac{d\underline{p}}{dt} = T(q_{a|bc})\underline{p}, \quad (8.3)$$

where the matrix $T(q_{a|bc})$ encapsulates the details of our chosen dynamics. The pair approximation, in which we discard correlations of third and higher order, lets us simplify this to

$$\frac{d\underline{p}}{dt} = T(q_{a|b})\underline{p}. \quad (8.4)$$

When people started doing simulations of lattice models like these, they found that the conditional probabilities $q_{a|M}$ equilibrate. That is to say, even if the global density of mutants p_M changes, the local statistical structure of a mutant cluster remains constant. This is the key statement which allows us to linearize the dynamics and write the behavior of $\underline{p}$ in terms of eigenvectors and eigenvalues:

$$\underline{p}(t) = C\underline{e}_A \exp(\lambda t). \quad (8.5)$$

The dominant eigenvalue λ of T is the “invasion exponent” which characterizes whether an invasion will fail ($\lambda < 0$) or succeed ($\lambda > 0$). The eigenvector $\underline{e}_A$ associated with λ describes the vehicle of selection for the mutants’ particular genetic variation, by summarizing the structure of their geographical cluster.

All of this, of course, is only as good as our linearization! If something interesting happens further away from the fixed point, just looking at the eigenvalues we got from

8.2 Example 1: Birth, Death, Movement

our matrix T won't tell us about it. In addition, if the actual dynamics of the system tend to form patterns which can't be represented very well by pairwise correlations, then pair approximation will run into a wall. As long ago as 1994, Tainaka [280] pointed out that, in a rock-paper-scissors system,

The failure of the mean-field theory and PA model implies that the long-range correlation is essentially important for the pattern formation.

Minus van Baalen puts the issue in the following way:

The extent to which pair-dynamics models are satisfactory depends on the goal of the modeler. As we have seen, these models do not capture all of the phenomena that can be observed in simulations of fully spatial probabilistic cellular automata. Basically, the approximation fails whenever spatial structures arise that are difficult to “describe” using pairs alone. More technically, the method fails whenever significant higher-order correlations arise – that is, whenever the frequency of paritucular triplets (or triangles, squares, or all sorts of star-like configurations) starts to diverge from what one would expect on the basis of pair densities. Thus, pair-dynamics models satisfactorily describe probabilistic cellular automata in which only “small-scale” patterns arise. Larger, “meso-scale” patterns such as spirals are difficult to capture using this method.

—in Dieckmann *et al.* (2000), chapter 19 [104].

It's also pretty easy for the algebra involved in a pair-approximation calculation to blow up far beyond the point of being useful. For example, Dobrinevski *et al.* [281] study a four-species system, where the pair approximation turns out to require 256 coupled differential equations. The only way to tackle that problem is to give it back to the computer and solve those equations numerically—and when they do that, it doesn't even work all that well!

8.2 Example 1: Birth, Death, Movement

In Chapter 3, and again in the previous section, we introduced the idea of *pair approximation*, by which we try to understand a system by tracking the joint probability distributions for pairs of its pieces. Now, we'll look at this machinery in more detail by focusing on a specific example. The ecosystem which we shall study will contain one species living on a regular lattice, and the individual organisms of that species can move about, give birth and die. That is, our pair dynamics will include three processes, each occurring stochastically with its own characteristic rate: *movement* or *migration*, *birth* and *death*. We follow the notation of van Baalen [104].

We write z for the “coordination number” of the lattice. That is, each lattice site will have z neighbors. We can represent the birth process as follows:



8 Invasion Fitness by Moment Closure Approximations

and we say this takes place with rate b/z .

Similarly, the death process can be represented as



for any site type a . This takes place with rate d/z .

Movement or migration from one site to a neighboring location is the reaction



This reaction occurs with rate m/z .

Given these reactions, we can write differential equations for the time derivatives of the pairwise densities. The density of $R0$ pairs changes as

$$\begin{aligned} \frac{dp_{R0}}{dt} = & -p_{R0}[b/z + d + (z-1)q_{0|R0}m/z + (z-1)q_{R|0R}(b+m)/z] \\ & + p_{00}(z-1)q_{R|00}(b+m)/z \\ & + p_{RR}[d + (z-1)q_{0|RR}m/z]. \end{aligned} \quad (8.9)$$

Likewise,

$$\frac{dp_{00}}{dt} = -p_{00}2(z-1)q_{R|00}(b+m)/z + p_{R0}2[d + (z-1)q_{0|R0}m/z]. \quad (8.10)$$

Finally,

$$\frac{dp_{RR}}{dt} = p_{R0}2[b/z + (z-1)q_{R|0R}(b+m)/z] - p_{RR}2[d + (z-1)q_{0|RR}m/z]. \quad (8.11)$$

Summing over pairwise densities recovers overall densities:

$$p_i = \sum_j p_{ij}. \quad (8.12)$$

From this, we deduce that

$$\frac{dp_R}{dt} = (bq_{0|R} - d)p_R. \quad (8.13)$$

If we ignore spatial structure altogether, we can say that

$$q_{0|R} = p_0, \quad (8.14)$$

which by normalization of probability means

$$q_{0|R} = 1 - p_R. \quad (8.15)$$

So,

$$\frac{dp_R}{dt} = (b(1 - p_R) - d)p_R. \quad (8.16)$$

This should look familiar: it's a logistic equation for population growth, with growth

rate $b - d$ and equilibrium population $1 - d/b$.

It's worth pausing a moment here and using this result to touch on a more general concern. Often, a logistic-growth model is presented with the growth rate and the equilibrium population size as its parameters. When we see the model in that form, we naturally start thinking of those parameters as *independently variable* quantities. We imagine that a mutation or a change in the environmental conditions could change one without affecting the other. However, if the growth rate and the equilibrium population size are both functions of other parameters taken together, then the changes *which are biologically reasonable to consider* will likely affect both of them. To understand which quantities we should treat as independent, we need to spend time looking at how the numbers which apply to population-scale phenomena arise from the smaller-scale physiological and ecological goings-on [282].

8.3 Example 2: Epidemic in an Adaptive Network

We can juice things up a little by considering another example, one which is topical for these jittery times. Let's look at *the spread of an epidemic*. There's a classic genre of models for this, which we can call after a prototypical representative, the Susceptible–Infected–Recovered or SIR model. In the SIR model, we imagine a population of individuals through which a disease can spread. Each individual is either **S**usceptible to the disease, **I**nfected with it or **R**ecovered from it.

Contact with an Infected individual can turn a Susceptible one into an Infected, so an S plus an I becomes an I plus an I . If the disease runs its course in an individual, they gain the status of R and are thereafter immune to further infection. (Perfect immunity is, mathematically, the same as death, but we'll be optimistic with our labels today.) We can complicate the model in many ways, for example by making the immune response *imperfect*, so that individuals who have recovered can be re-infected later. This could happen by the immunity fading over time, so that R individuals transition back to S , or the immunity might only be partial, so that we have transitions from R directly back to I . We can also add *population structure*: interesting things happen when the individuals are not all in direct contact with one another. This complication is obviously something we'll have to address if we want to do epidemiology with real-world diseases! Speaking from a more mathematical perspective, we find neat phase-transition effects when we put these epidemic models on a lattice; see Chapter 3 and references therein.

The complicating factor we'll consider in this section is the following: the spread of a disease through a social network can itself change the way people contact each other. This makes epidemiology a candidate subject for the study of adaptive networks: graph-structured systems in which the states associated with the vertices and the topology of the edges can change on the same timescales, feeding back on one another [67, 68, 70, 71, 110, 167, 168, 169, 171, 283].

To the SIR problem we described earlier, we add two extra wrinkles: first, the individuals are arranged in a network, and infection spreads only along the links within that network. Second, if a susceptible individual is in contact with an infected one,

8 Invasion Fitness by Moment Closure Approximations

that link can be broken, and a new link established to another susceptible individual, which we pick at random from the pool of eligible susceptibles (that is, those who aren't already neighbours—we are disallowing multiple links). This *rewiring rule* makes this scenario an adaptive-network problem.

For the moment, let's say that once they're infected, these organisms don't recover. We have only S and I in the population at any time. Therefore, the probability that an organism chosen at random has status S and the probability that an organism chosen at random has status I sum to 1:

$$p_S + p_I = 1. \quad (8.17)$$

This simplifies the SIR model to an SIS model. A node can start in the S state, become infected and enter the I state, and then potentially recover and return to the S state again. We neglect the possibility of immunity: nodes which have been infected and recovered are just as susceptible to the disease as those which have never been in the I state.

Following the literature [110, 167], we choose to normalize our pairwise densities in the following way:

$$p_{SS} + p_{SI} + p_{II} = \langle k \rangle, \quad (8.18)$$

where $\langle k \rangle$ is the average degree of the nodes in the network.

The number of infected individuals decreases as organisms recover from the disease, while it increases as the contagion spreads over links from those already infected. We say how quickly recovery happens using the parameter r , and we encode the ease with which the disease travels along network links with the transmissibility τ . With these definitions, we can write the following rate equation for the density of infected individuals, p_I :

$$\frac{d}{dt} p_I = \tau p_{SI} - r p_I. \quad (8.19)$$

In the original SIR example, where everyone was in contact with everyone else, we could say $p_{SI} \approx \langle k \rangle p_S p_I$. But we ought to be wary of using this approximation here, for two reasons. First, any time we have a system which has a chance of developing heterogeneity, of forming lumps in one region which don't directly affect lumps in another, then averaging over the whole system becomes a risky business. Second, more specifically, this approximation can't capture rewiring. Links are breaking and re-forming all the time in our model, but the product $p_S p_I$ stays the same when we move links around. So, to write an equation for how p_I changes, we need to write down how the density of SI pairs will change, but in order to do that, we have to include the rewiring effect.

Let's introduce a third parameter, w , to indicate how much rewiring is going on. The w parameter will be a rate, having units of inverse time, just like r and τ . The density of SS pairs will go *down* as the disease spreads to them from infected nodes, but it will go *up* as nodes recover *and* as susceptible nodes rewire their links. Consequently, the time derivative of p_{SS} must include a positive term which depends on r and w ,

8.3 Example 2: Epidemic in an Adaptive Network

and a negative term which depends upon τ .

$$\frac{d}{dt}p_{SS} = (r + w)p_{SI} - \tau p_{SSI}. \quad (8.20)$$

Similarly, p_{II} is increased by the disease being transmitted, and decreased by the rewiring effect:

$$\frac{d}{dt}p_{II} = \tau(p_{SI} + p_{ISI}) - 2rp_{II}. \quad (8.21)$$

Now comes the moment-closure step. We enforce a pair approximation, by writing three-way probabilities in terms of lower-order ones:

$$p_{ISI} = \frac{\langle q \rangle}{\langle k \rangle} \cdot \frac{p_{SI}^2}{p_S}. \quad (8.22)$$

The “mean excess degree” $\langle q \rangle$ is the number of additional links we expect to find after we follow a random link. It turns out [110] that $\langle q \rangle = \langle k \rangle$ is a reasonable approximation.

Our dynamical system is defined by three equations. First, we have the rule we wrote before,

$$\frac{d}{dt}p_I = \tau p_{SI} - rp_I, \quad (8.23)$$

and then the rules we deduced using pair approximation:

$$\frac{d}{dt}p_{SS} = (r + w)p_{SI} - 2\tau p_{SI} \frac{p_{SS}}{p_S}, \quad (8.24)$$

and

$$\frac{d}{dt}p_{II} = \tau p_{SI} \left(1 + \frac{p_{SI}}{p_S} \right) - 2rp_{II}. \quad (8.25)$$

Having written these equations, we can compare the behaviour of the dynamical system they define to that of a simulation of the original model. They actually agree pretty well [110, 167].

Note that we have made an important simplifying assumption, not in the analytical treatment of our model but in the original definition of it: we chose a rewiring rule which lets nodes form new links to S -type nodes anywhere else in the system. That is, the rewiring rule lacks any notion of *proximity*. Depending on the phenomenon we’re trying to model, it might be reasonable to allow rewiring only to neighbours of a node’s current neighbours, for example. This could be the case if information about available opportunities for building new connections had to spread through existing links. Or, if the nodes’ spatial positions are significant, perhaps links which span longer geographical distances should be disfavoured. These are the sorts of complications which can make pair approximations inapplicable.

The study of dynamical processes on and of graph structures brings together the subjects of “network theory”, as the term has been used in stochastic process research, and “complex networks”, to invoke a term which has been trending elsewhere.

8.4 Example 3: Evolution of Altruism

One application of this machinery has been to understand how evolution can produce *altruistic behaviour* [103]. A behaviouristic definition of “altruism” would go something like, “Acting to increase the reproductive success of another individual at the expense of one’s own”. This can be written out using game theory; one defines a parameter b to stand for benefit, c to stand for cost and a payoff function which depends on b , c and the strategy employed by an organism.

We shall consider a lattice of sites, each one of which can be in one of three states: *empty*, denoted by 0; occupied by a *selfish organism*, denoted by S ; and occupied by an *altruist*, denoted by A .

In the reproduction process, an organism spawns an offspring into an adjacent empty lattice site. This turns a pair of type $S0$ into a pair of type SS . At what rate should the transition $S0 \rightarrow SS$ occur? If we presume some baseline reproductive rate, call it b_0 , then the presence of altruistic neighbours should *augment* that rate. We’ll say that if the number of nearby altruists is n_A , then selfish individuals will reproduce at a rate $b_0 + Bn_A/n$, where the parameter B specifies how helpful altruists are. The reproduction process for altruists, which we can write $A0 \rightarrow AA$, occurs at a rate $b_0 + Bn_A/n - C$. Here, the parameter C is the *cost* of altruism: it’s how much an altruist gives up to help others.

In the differential equation for dp_{S0}/dt , the $S0 \rightarrow SS$ transition contributes a term proportional to the density of $S0$ pairs:

$$-[(1 - \phi)b_S + (b_S + m)\phi q_{S|0S}]p_{S0}, \quad (8.26)$$

where we have written ϕ for $1 - 1/z$, to save a little ink. All things told, the rate of change of p_{S0} is given by

$$\begin{aligned} \frac{dp_{S0}}{dt} = & (b_S + m)\phi q_{S|00}p_{00} \\ & - [d_S + m\phi q_{0|S0}]p_{S0} \\ & - [(1 - \phi)b_S + (b_S + m)\phi q_{S|0S}]p_{S0} \\ & + [d_S + m\phi q_{0|SS}]p_{SS} \\ & - (b_A + m)\phi q_{A|0S}p_{S0} \\ & + [d_A + m\phi q_{0|AS}]p_{SA}. \end{aligned} \quad (8.27)$$

Yuck! After writing a few equations like that, it’s easy to wonder if maybe we should look for *new mathematical ideas* which could help us better organise our thinking. But, for the nonce, we will simply press on with the algebra.

The next steps follow the general plan we laid out above. We write differential equations for the pairwise probabilities p_{ab} , which depend on triplet quantities $q_{a|bc}$. Then, we impose a pair approximation, declaring that $q_{a|bc} = q_{a|b}$, which gives us a closed system of equations. Next, we find the fixed point with $p_A = 0$, and we perturb around that fixed point to see what happens when a strain of altruists is introduced into a selfish population. The dominant eigenvalue λ of the time-evolution matrix T tells us, in this approximation, whether the altruistic strain will invade the lattice or

8.4 Example 3: Evolution of Altruism

wither away. The condition $\lambda > 0$ can be written in the form

$$B\phi\tilde{q}_{A|A} - C > 0. \quad (8.28)$$

Here, we've written $\tilde{q}_{A|A}$ for the conditional probability of altruists contacting altruists which obtains as the local densities equilibrate. That is to say, an attempted invasion by altruists will succeed if a measure of benefit, B , multiplied by an indicator of “assortment” among genetically similar individuals, is greater than the cost of altruistic behaviour, C .

After all our mucking with eigenvalues, we have found a condition which is strongly reminiscent of a classic and influential idea from mid-twentieth-century evolutionary theory. In biology, the inequality

$$(\text{benefit}) \times (\text{relatedness}) > (\text{cost}) \quad (8.29)$$

is known as *Hamilton's rule* [55]. This is a rule-of-thumb for when natural selection can favour altruistic behaviour: altruists can prosper when the inequality is satisfied. Hamilton's rule was originally derived for unstructured populations, with no network topology or spatial arrangement to them. We can understand Hamilton's rule in this context in the following way:

How well an organism fares in the great contest of life depends on the environment it experiences. During the course of its life, an individual member of a species will interact with a set of others, which we could call its “social circle”. The composition of that social circle affects how well an individual will propagate its genetic information to the next generation—its *fitness*. In an unstructured population, we can think of such circles being formed by taking random samples of the population. An altruist, by our definition, sacrifices some of its own potential so that offspring of other individuals can prosper. A social circle of altruists can fare better than a social circle of selfish individuals, increasing the chances that social circles which form in the next generation will contain altruists [55].

It's common to treat “benefit” and “cost” as parameters of the system. We could potentially derive them from more fundamental dynamics, if we looked more closely at the interactions within a particular ecosystem, but right now, they're just knobs we can turn. What about the remaining quantity in Hamilton's rule: what does “relatedness” mean? Excellent question! We can get a feel for where the term came from by taking a gene's-eye view: copies of many of my particular genetic variants will be sitting inside the cells of my close relatives. Consequently, as far as my genes are concerned, if my relatives survive, that's almost as good as my surviving. When reckoning the benefit of altruism against its cost, then, the aid one organism brings to another ought to be weighted by how “related” they are.

So, we can say that we have “recovered Hamilton's rule as an emergent property of the spatial dynamics”—*if* we are willing to draw a circle around the middle of our formula and declare those terms to be the “relatedness”.

Knowing where our invasion condition came from, we can appreciate some of the caveats which scientists have raised in connection with Hamilton's rule.

Simon *et al.* address this specifically [153]:

In particular, r is often taken to be the average relatedness of interacting individuals, as compared to the average relatedness in the population, in which case inequality (1) [$rB > C$] is referred to as Hamilton’s rule. It is important to note that inequality (1) is only a description of whether the current level of assortment as subsumed in the parameter r is sufficient to favour cooperation, but not a description of the mechanisms that would lead to such assortment. It has been suggested repeatedly that the problem of cooperation can be understood entirely based on Hamilton’s rules of the form (1). Even though often taken as gospel, this claim is wrong in general, for two reasons.

First, and foremost, even if a rule of the form (1) predicts the direction of selection for cooperation at a given point in time, the long-term evolution of cooperation cannot be understood without having a dynamic equation for the quantity r , i.e., without understanding the temporal dynamics of assortment. The dynamics of r in turn cannot be understood based solely on the current level of cooperation, and hence expressions of type (1) are in general insufficient to describe the evolutionary dynamics of cooperation. Second, the quantity r , which measures the average relatedness among interacting individuals, is insufficient to construct Hamilton’s rule in models that account for variable individual-level death rates and/or group-level events.

Damore and Gore [47] have more to say on this point:

Contrary to the popular use of the word, “relatedness” describes a population of interacting individuals, where r refers to how assorted similar individuals are in the population.

And in further detail:

[E]very definition of relatedness must take into account the population. Therefore, relatedness is not the percent of genome shared, genetic distance, or any extent of similarity between two isolated individuals in a larger population. Also, because horizontal gene transfer is commonplace between microbes and selection is strong, phylogenetic distance or any other indirect genetic measure is likely to be inaccurate. Many of these false definitions live on partly because ambiguous heuristics like “ $\frac{1}{2}$ for brothers, $\frac{1}{8}$ for cousins,” which require very specific assumptions, are repeated in the primary literature. Also, most non-theoretical papers simply define relatedness as “a measure of genetic similarity” and do not elaborate or instead leave the precise definition to the supplemental information [...]. Unfortunately, scientists can easily misinterpret this “measure of genetic similarity” to be anything that is empirically convenient such as genetic distance or percent of genome shared. Largely because of this confusion, we support the more widespread use of the term “assortment,” which is

8.4 Example 3: Evolution of Altruism

harder to misinterpret [...] For similar reasons of reader understanding, we also encourage authors to make calculations more explicit, either in the main or supplemental text, and to avoid repeating previous results without giving the assumptions that went into deriving them.

It is for this reason that we called $\phi\tilde{q}_{A|A}$ a measure of “assortment” earlier. Of course, even with this careful choice of terminology, the limitations of our Hamilton-esque rule still apply: we know that because we derived it from the condition that the dominant eigenvalue be positive, it will miss any effects which a fixed-point eigenvalue analysis is not sensitive to.

Stepping back for a moment, notice that although the terms and coefficients started to proliferate on us, we haven’t introduced any remarkably “advanced” or “esoteric” mathematics. Derivatives, matrices, eigenvalues—this is *undergraduate* stuff! The amount of algebra we’ve been able to stir up without really even trying is, however, a little worrying. We can *invent* a mathematical model for some particular biological scenario, and we might even be able to solve it, or at least tell how it’ll behave in certain interesting circumstances. But what if we want *general* results which extend across models, or ideas which will help us identify the common features and the key disparities among a host of examples? We will return to this question in Chapter 10.

We conclude this section with a more detailed derivation of the invasion condition (8.28). First, for convenience’s sake, we define the abbreviations

$$\alpha_j = (b_j + m)\phi = (b_j + m)(1 - 1/z), \quad (8.30)$$

$$\beta_j = (1 - \phi)b_j = \frac{b_j}{z}, \quad (8.31)$$

$$\gamma_{jk} = d_j + m\phi q_{0|jk}. \quad (8.32)$$

These denote, respectively, the effective rates of arrivals from neighboring sites, birth events within pairs of sites, and the vacation of sites by migration or death. If there are no altruists in the ecosystem, then the $S0$ and SS pair densities evolve according to the coupled equations

$$\frac{dp_{S0}}{dt} = \alpha_S q_{S|00} p_{00} - (\beta_S + \alpha_S q_{S|0S} + \gamma_{S0}) p_{S0} + \gamma_{SS} p_{SS}, \quad (8.33)$$

$$\frac{dp_{SS}}{dt} = 2(\beta_S + \alpha_S q_{S|0S}) p_{S0} - 2\gamma_{SS} p_{SS}. \quad (8.34)$$

Setting these rates to zero determines the equilibrium point for the altruist-less ecosystem. At this equilibrium,

$$\alpha_S \bar{q}_{S|0S} \bar{p}_{00} - \gamma_{S0} \bar{p}_{S0} = 0. \quad (8.35)$$

Invoking the pair approximation, we drop some indices:

$$\alpha_S \bar{q}_{S|0} \bar{p}_{00} - \gamma_S \bar{p}_{S0} = 0, \text{ where } \gamma_S = d_S + m\phi q_{0|S}. \quad (8.36)$$

8 Invasion Fitness by Moment Closure Approximations

Recalling that

$$\bar{q}_{S|0} = \frac{\bar{p}_{S0}}{\bar{p}_0}, \quad \bar{q}_{0|0} = 1 - \bar{q}_{S|0} = \frac{\bar{p}_{00}}{\bar{p}_0}, \quad (8.37)$$

we conclude that

$$\alpha_S \bar{q}_{0|0} - \gamma_S = 0. \quad (8.38)$$

This tells us the density of adjacent empty sites in an equilibrated S -type population.

Next, we consider a resident population of selfish individuals that has come to an equilibrium and is invaded by a strain of altruists. Will the invasion be successful or not? We make the approximation that initially, the altruists are sufficiently rare that they do not affect the resident population, so the demographic statistics we derived above are still applicable.

The pairwise densities involving altruists then evolve according to the following three coupled differential equations:

$$\frac{dp_{A0}}{dt} = \alpha_A q_{A|00} \bar{p}_{00} - (\beta_A + \alpha_A q_{A|0A}) p_{A0} - \alpha_S q_{S|0A} p_{A0} + \gamma_{SA} p_{AS} - \gamma_{A0} p_{A0}, \quad (8.39)$$

$$\frac{dp_{A0}}{dt} = \alpha_A q_{A|0S} \bar{p}_{0S} + \alpha_S q_{S|0A} p_{A0} - (\gamma_{AS} + \gamma_{SA}) p_{AS}, \quad (8.40)$$

$$\frac{dp_{A0}}{dt} = 2(\beta_A + \alpha_A q_{A|0A}) p_{A0} - 2\gamma_{AA} p_{AA}. \quad (8.41)$$

We are treating the barred quantities as constant, so the dynamical variables are p_{A0} , p_{AS} and p_{AA} . Collecting these into a three-element vector $(p_{A0}, p_{AS}, p_{AA})^T$, we can express this dynamical system as

$$\frac{d\vec{p}}{dt} = M(q_{j|kl}) \vec{p}, \quad (8.42)$$

where the matrix M is a function of the triplet densities:

$$M(q_{j|kl}) = \begin{pmatrix} \alpha_A q_{A|00} - (\beta_A + \alpha_A q_{A|0A}) - \gamma_{A0} - \alpha_S q_{S|0A} & \gamma_{AS} & \gamma_{AA} \\ (\alpha_S + \alpha_A) q_{S|0A} & -\gamma_{AS} - \gamma_{SA} & 0 \\ 2(\beta_A + \alpha_A q_{A|0A}) & 0 & -2\gamma_{AA} \end{pmatrix}. \quad (8.43)$$

Here, we have used the fact that

$$q_{A|0S} \bar{p}_{0S} = q_{S|0A} p_{0A} = p_{A0S} \quad (8.44)$$

to simplify the middle element in the left column.

Again calling upon the pair approximation, we simplify the matrix M to

$$M(q_{j|k}) = \begin{pmatrix} \alpha_A - \beta_A - (\alpha_A + \alpha_S) \bar{q}_{S|0} & \gamma_S & \gamma_A \\ (\alpha_S + \alpha_A) \bar{q}_{S|0} & -\gamma_A - \gamma_S & 0 \\ 2\beta_A & 0 & -2\gamma_A \end{pmatrix}. \quad (8.45)$$

We are now in the familiar terrain of linear stability analysis. The altruist invasion

will be successful, according to all the approximations we have made so far, if the all- S fixed point is unstable. The transition between stability and instability occurs when the dominant eigenvalue of M changes from negative to positive. The product of the eigenvalues of M is the determinant of M , so we seek the location in parameter space where the determinant of M vanishes. This yields the condition

$$(\alpha_A - \gamma_A)(\gamma_A + \gamma_S) - (\alpha_A + \alpha_S)\gamma_A\bar{q}_{S|0} = 0. \quad (8.46)$$

Having derived $\bar{q}_{0|0}$ in Eq. (8.38), we have a value we can use for $\bar{q}_{S|0}$. Making this substitution,

$$(\alpha_S + \gamma_A)(\alpha_A\gamma_S - \alpha_S\gamma_A) = 0, \quad (8.47)$$

meaning that the transition lies at

$$\frac{\alpha_A}{\gamma_A} = \frac{\alpha_S}{\gamma_S}. \quad (8.48)$$

The all- S fixed point is unstable when the ratio α_A/γ_A exceeds α_S/γ_S . This provides our condition for invasion success:

$$\frac{(b_A + m)\phi}{d_A + m\phi\tilde{q}_{0|A}} > \frac{(b_S + m)\phi}{d_S + m\phi\tilde{q}_{0|S}}, \quad (8.49)$$

where $\tilde{q}_{0|j}$ are given by the eigenvector of M corresponding to the dominant eigenvalue.

To simplify still further, let's see what happens when the migration rate goes to zero, and the death rates d_A and d_S are equal. Then the invasion condition reduces to $b_A > b_S$, or

$$\boxed{B\phi\tilde{q}_{A|A} - C > 0.} \quad (8.50)$$

This has the form we promised: the altruists invade if the benefit parameter, multiplied by an assortment factor that depends on population structure, is greater than the cost parameter. Note that the assortment factor will generally depend upon the population dynamics, through quantities like migration rates!

8.5 Host–Consumer Dynamics

Let us consider a spatial host–consumer system containing a single type of consumer. In such an ecosystem, each lattice site can be in one of three possible states: 0, H and C . Consequently, there are six different site-site correlations. Because

$$\sum_j P(ij) = P(i), \quad (8.51)$$

only three of the six variables are independent. Following de Aguiar *et al.*, we choose the independent correlations to be

$$u = P(H0), \quad r = P(HC) \text{ and } w = P(0C). \quad (8.52)$$

8 Invasion Fitness by Moment Closure Approximations

That is, our independent variables will be x , y , u , r and w . This choice implies that if we write $z = 1 - x - y$ for the fraction of empty space, the other three correlations are given by

$$q = P(00) = z - u - w, \quad p = P(HH) = x - r - u \text{ and } s = P(CC) = y - r - w. \quad (8.53)$$

A given host can only be consumed once. Therefore, the simplest way to find the probability of its consumption is to calculate one *minus* the probability that it is *not* consumed. If a host has m consumer neighbors, then this probability is $1 - (1 - \tau)^m$. Analogous considerations apply to empty sites being colonized by hosts.

Let ζ be the number of nearest-neighbor sites (so, on the default square lattice, $\zeta = 4$). We define

$$h_\zeta(\alpha) = 1 - (1 - \alpha)^\zeta. \quad (8.54)$$

The host population fraction x increases if hosts can reproduce into empty space, which depends upon the contact rate between host sites and empty sites. Therefore, dx/dt will have a positive contribution that depends on u . It will also have a negative contribution that depends upon the probability that consumers will be adjacent to hosts and thus able to consume them. Together, these two processes combine to yield

$$\frac{dx}{dt} = z h_\zeta(gu/z) - x h_\zeta(\tau r/x). \quad (8.55)$$

The rate of change in y can be written similarly. However, consumers do not require the presence of any particular entities in their neighborhood to die off, so the negative term will only depend on y itself:

$$\frac{dy}{dt} = x h_\zeta(\tau r/x) - vy. \quad (8.56)$$

If we make the approximation $r = xy$, this becomes

$$\frac{dy}{dt} = x h_\zeta(\tau y) - vy, \quad (8.57)$$

which is the mean-field equation we would write for consumer dynamics.

On much the same principles, we can write time-evolution equations for the pairwise correlation variables. The appropriate equations turn out to be

$$\begin{aligned} \frac{du}{dt} &= (q - u) h_{\zeta-1}(gu/z) + vr - u h_{\zeta-1}(\tau r/x) \\ &\quad - gu[1 - h_{\zeta-1}(gu/z)], \end{aligned} \quad (8.58)$$

and

$$\begin{aligned} \frac{dr}{dt} &= (p - r) h_{\zeta-1}(\tau r/x) - vr + w h_{\zeta-1}(gu/z) \\ &\quad - \tau r[1 - h_{\zeta-1}(\tau r/x)] \end{aligned} \quad (8.59)$$

and

$$\frac{dw}{dt} = u h_{\zeta-1}(\tau r/x) + v(s-w) - w h_{\zeta-1}(gu/z). \quad (8.60)$$

This is the system of differential equations we solved numerically in order to find the pair-approximation results in Chapter 3. They can be extended to include two types of consumer with differing τ values. However, investigating that augmented system of equations reveals a problem [95, 100, 101]: higher τ almost always wins out over lower τ . This means that, as we saw in Chapter 3, whatever we learn from pair approximation must be augmented by knowledge from another source if we wish to understand evolutionary processes.

8.6 Modified Mean-Field

We should at this point mention a proposal by Pascual *et al.* [284] to incorporate spatial effects into mean-field models by altering the functional dependence on mean-field quantities in differential equations, rather than by moment closures. Pascual *et al.* begin with a simple Lotka–Volterra model for a plant species, whose population density they denote p , and an herbivore species, whose density is denoted h :

$$\dot{p} = \alpha p e - \beta p h, \quad (8.61)$$

$$\dot{h} = \beta p h - \delta h. \quad (8.62)$$

Here, e stands for the density of empty space, $1 - (p + h)$. By rescaling the units of time, we can eliminate one parameter, which we pick to be the consumption rate β :

$$\dot{p} = \alpha p e - p h, \quad (8.63)$$

$$\dot{h} = p h - \delta h. \quad (8.64)$$

This system of equations has a stable equilibrium point at

$$p^* = \delta, \quad h^* = \frac{\alpha(1-\delta)}{1+\alpha}. \quad (8.65)$$

Pascual *et al.* suggest modifying the dependence of $\dot{p}$ and $\dot{h}$ on the variables p , e and h , so that each variable no longer enters linearly:

$$\dot{p} = \alpha c_1 p^{a_1} e^{b_1} - c_2 p^{a_2} h^{b_2}, \quad (8.66)$$

$$\dot{h} = c_2 p^{a_2} h^{b_2} - \delta h. \quad (8.67)$$

This is the “modified mean-field” (MMF) model. The new parameters a_i and b_i are intended to represent the deviations in the effective population density from mean-field expectations which an individual organism experiences due to spatial fluctuations. The other new parameters, c_1 and c_2 , incorporate linear effects of spatial pattern formation which “can decrease or increase (when $0 < c_i < 1$ or $c_i > 1$ respectively) the overall

rate but cannot not modify its functional form” [284]. The MMF may appear to be a rather ad-hoc alteration of the original Lotka–Volterra system, but the equations are simpler than those produced by pair approximation, and by fitting the new parameters, the trajectories of p and h seen in a lattice system can be reproduced.

For our purposes, however, we need more. Can the MMF say anything about the *competition* between two types of herbivore with different consumption rates? This is the minimum we require when we advance from ecological to eco-evolutionary dynamics. To investigate this, we consider an augmented MMF system with population densities p , h and k . Now, the amount of empty space is given by

$$e = 1 - (p + h + k). \quad (8.68)$$

The dynamical equations are

$$\dot{p} = \alpha c_1 p^{a_1} e^{b_1} - c_2 p^{a_2} h^{b_2} - c_3 p^{a_2} k^{b_2}, \quad (8.69)$$

$$\dot{h} = c_2 p^{a_2} h^{b_2} - \delta h, \quad (8.70)$$

$$\dot{k} = c_3 p^{a_2} k^{b_2} - \delta k. \quad (8.71)$$

The new parameter c_3 indicates the consumption rate of the additional herbivore species, which we can think of as a mutant variety. Without loss of generality, we take $c_3 > c_2$. If $k = 0$, we recover the two-species MMF system. Moreover, the only difference between the two types of herbivore is the rate at which they consume plants, so we have no reason to think that the exponents should be different (and in fact Pascual *et al.* find no significant dependence of the fitted exponents a_2 and b_2 on the rate parameters).

From the difference

$$\frac{d}{dt}(h - k) = \dot{h} - \dot{k} = p^{a_2}(c_2 h^{b_2} - c_3 k^{b_2}) - \delta(h - k), \quad (8.72)$$

we see that if $h = k$, then the right-hand side is negative, and so $\dot{h} < \dot{k}$. Consequently, if the resident and mutant herbivore population densities are ever equal, then the mutant population is growing faster, or is diminishing less rapidly. This implies that the resident population density cannot cross from below the mutant density to above. Invasion by a strain with lower voraciousness is, in the MMF, impossible. However, we see it happening easily in the spatial host–consumer model.

Numerical analysis (Figure 8.1) shows that a small amount of a mutant herbivore strain with higher voracity can take over a two-species system which has settled into its equilibrium. In other words, the MMF approach isn’t any better at handling coexistence or “the resident strikes back” effects than moment closure is [100, 101]. This is only to be expected, since neither approach captures the descendent-shading we examined in Chapter 3. One can hack the exponents until the MMF model matches the two-species ecological dynamics, but it is inadequate for *evolutionary* studies.

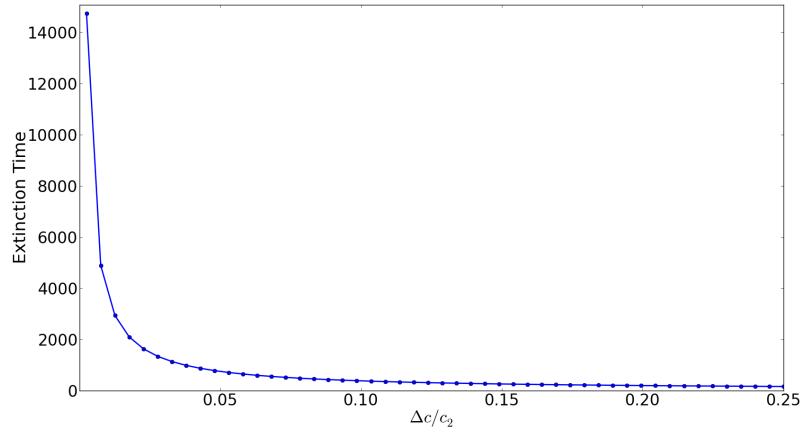


Figure 8.1: Time required for a mutant herbivore population to take over an ecosystem, as a function of the difference in consumption rates, $\Delta c = c_3 - c_2$. The mutant population density is initialized to 1/100 of the resident, and the dynamics are simulated until the resident population density has diminished to 1/100 of its initial value. To facilitate comparison with the original paper [284], parameter values for these computations were set at $a_1 = 0.2$, $b_1 = 0.59$, $c_1 = 0.19$, $a_2 = 0.59$, $b_2 = 1.0$, $c_2 = 0.4$, $\alpha = 0.6$, $\delta = 0.25$.

9 The Varieties of Multilevel Selection

9.1 Introduction

Ever since Darwin, biologists have tried to explain how social behaviors—cooperation, even self-sacrifice—could emerge from natural selection, which seems to embody a dog-eat-dog, dog-eat-cat world. One explanation for how evolution could yield altruism, popular up through the 1950s, went something like this: imagine a species divided into several smaller sub-populations, like flocks, tribes or wandering herds. Some of the herds might be full of selfish individuals who act without regard to others in their vicinity, but other sub-populations might be full of altruists, whose cooperative behavior allows those sub-populations to produce more sub-populations, also full of altruists. If something in the environment kills off sub-populations, then the clusters of altruists will produce more offspring clusters than those comprised of selfish individuals, and this biased survival of stochastically varying replicators—in this case, the replicating units being taken as clusters—will lead to a species dominated by altruists.

In fact, this idea goes back to Darwin [285], who wrote,

Natural selection will modify the structure of the young in relation to the parent, and of the parent in relation to the young. In social animals it will adapt the structure of each individual for the benefit of the whole community; if the community profits by the selected change.

And, in more detail [286]:

When two tribes of primeval man, living in the same country, came into competition, if the one tribe included (other circumstances being equal) a greater number of courageous, sympathetic, and faithful members, who were always ready to warn each other of danger, to aid and defend each other, this tribe would without doubt succeed best and conquer the other. [...] Selfish and contentious people will not cohere, and without coherence nothing can be effected. A tribe possessing the above qualities in a high degree would spread and be victorious over other tribes; but in the course of time it would, judging from all past history, be in its turn overcome by some other and still more highly endowed tribe. Thus the social and moral qualities would tend slowly to advance and be diffused throughout the world.

However, as Darwin recognized, this process might be vulnerable to subversion from within:

9 The Varieties of Multilevel Selection

It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those which were the most faithful to their comrades, would be reared in greater number than the children of selfish and treacherous parents of the same tribe.

Understanding this problem more deeply requires a better knowledge of the mechanisms of heredity than was available in Darwin’s time. Indeed, genetics did not become a serious part of evolutionary theorizing until the early twentieth century [287]. Now we know that each human being has far more cells in their brain than genes in their genome, so we must expect that genes code for *patterns of development*. In turn, the structures formed by those developmental patterns then display ranges of behavior, in ways which depend on the environment. And, as for our species, so for the others. A justifiably popular aphorism observes, “Evolution is the control of development by ecology” [288]. When we perform mathematical modeling, we often elide these intermediate steps and pretend that genetic information can specify behaviors directly. How much we lose by doing this is difficult to say in advance—but even the bower birds we modeled back in Chapter 2 turn out to have culture [289].

We can refer to the idea that natural selection acts upon groups of organisms as *group selection*. The general notion of natural selection effecting change at different scales within a population is *multilevel selection*. This has been a remarkably controversial idea [290, 291, 292], with a significant fraction of the controversy arising from people arguing words instead of mathematics. And, when mathematics was employed, it turned out to be too much statistics and too little dynamics [150, 151].

In order to sort this out, we have to be careful and delineate the various different ideas which have been paraded under the “multilevel selection” banner. All too often, the terminology of the subject has brought the appearance of precision, more than the actuality.

A useful parallel can be drawn between the different types of multilevel selection and the hierarchy of approximations used in spatial ecology. This relationship clarifies which modeling methods are equivalent, and it points the way to future extensions of multilevel selection theory. The key question is one of *description*: how much and what kind of information does a modeling method use to represent the state of being of an ecosystem? The different types of multilevel selection (MLS) answer this in different ways and thus relate to different kinds of ecological model. We shall see that “MLS-A” stands on a level with mean field theory, and “MLS-B” is akin to pair approximation.

Emphasizing the idea that context matters takes us beyond the mean field approximation. This extension becomes necessary if our model includes explicit group-level processes like budding and fusion, or if the geographical extent of our ecosystem fosters spatial heterogeneity. The much-discussed equivalence of multilevel selection and inclusive fitness (IF) is easy to establish for MLS-A and can be proven mathematically in the mean field, but real differences arise when we move from MLS-A to MLS-B and beyond.

The mathematical study of biological evolution, as it stands today, includes a wide variety of particular models but lacks an overall theory to organize those models and clarify their interrelationships [33]. This provides an opportunity for interdisciplinary

collaboration. Mathematicians and physicists can bring not just new techniques but also new *attitudes*, potentially creating a novel way of thinking about generalizations, specializations, equivalences and other ways mathematical models are interrelated. The challenge for the physicist is to know enough biology that one does not waste time on an obsolete or irrelevant problem, while also avoiding absorption in the internal politics and factionalism of long-running biologist debates. Interdisciplinary research must steer a course between the Scylla of diletantism and the Charybdis of tedium.

Simon *et al.* [153] characterize two types of MLS models, which we shall designate *MLS-A* and *MLS-B*. Their description of *MLS-A* states,

Key features typically include that all groups form at the same time from a randomly mixed (or mated) global population and contain two types (e.g. cooperators and defectors), the number of groups is constant or infinite, groups start out all the same size, groups vary only in the reproduction rate of the individuals, and all groups cease to exist at the same time. Groups then contribute their individuals to the random global mixing (or mating) phase from which new groups are again formed.

In *MLS-A*, the “groups” are social circles of organisms which form at random, typically once each generation. This random formation and reformation implies that the only information used to predict the environment experienced by any organism are *aggregate* measures taken over the entire population. Any single organism interacts with its local environment, *i.e.*, the randomly-formed group of which it is a part, but that local environment is modeled using a global average. This coarse-graining of the ecological picture is a *mean-field approximation* [48]. (We recall from Chapter 3 that this term, used in spatial ecology [95, 108] and evolutionary dynamics [68], derives originally from statistical physics [49].) Additional complications, such as nonlinear fitness functions, can make the mathematical analysis more cumbersome, but as long as the model uses only global averages to stand in for the ecosystem’s configuration, it remains within the mean field.

By contrast, *MLS-B* models “contain more explicit group-level events” such as group fissioning, group merging, mass dispersal and so forth [153]. An *MLS-B* model does not treat all configurations with the same total numbers of cooperators and defectors as equivalent. The allocation of individuals into groups matters, and most importantly, the composition of each group is not set by global averages [153, 154]. An organism’s local environment is no longer assumed to be representative of the aggregate properties of the whole population, and vice versa.

This advance is, in concept if not in algebraic detail, essentially the same as the move in spatial ecology from mean-field theory to higher-order *moment closures* such as *pair approximation*. We can illustrate this parallel by example: let us consider how one would treat a population of cooperators and defectors, arranged in a spatial or a network structure, by a pair approximation method. Because the population is structured, rather than being well-mixed, having a clump of cooperators in one small region is a significantly different circumstance than having the same number of cooperators spread uniformly throughout the whole ecosystem. To attempt to capture this distinction, we augment the bare-bones, coarse-grained description of

the ecosystem, moving beyond overall averages and including some measure of how local environments can differ. Consider the probability that an organism chosen at random from the population is a cooperator. This is an overall, global average; we can denote it p_C . Its complement, the probability that a randomly-picked organism is a defector, is $p_D = 1 - p_C$. (For simplicity, we assume that the region we are studying is entirely filled up, with no empty spaces.) If all the cooperators are clustered together in a patch, then the *conditional* probability that the *neighbor of a cooperator* is also a cooperator will be significantly different from the overall density of cooperators in the population at large. If we pick an organism at random and find that it is a cooperator, we should be willing to bet more heavily on the chance that picking one of its neighbors at random will find another cooperator. In short, the conditional probability $q_{C|C}$ can differ from p_C . A pair approximation for this model assumes that these pairwise conditional probabilities capture enough of the ecosystem’s heterogeneity to make all the predictions in which we are interested.

The move from mean-field theory to a pair approximation, like the move from MLS-A to MLS-B, brings in the idea that variation among local environments matters. Furthermore, both MLS-B and pair approximation share an important simplifying assumption of their own: variation among locales is important, but *the set of all locales has no structure*. Consider, for example, the MLS-B model of Simon *et al.* [153]. This model represents the state of the entire population at time t by a list of numbers, namely, the number of groups of type $\vec{x}$ at that time t , for all possible group-types $\vec{x}$. There is no sense in which some groups are geographically closer to each other than others. A group of type $\vec{x}$ in one spot is just as good as a group of type $\vec{x}$ in any other spot. Likewise, when we develop a pair approximation, a pair of a certain type (for example, cooperator–defector) occurring in one location counts the same as a pair of that type occurring anywhere else.

As we mentioned earlier, the parallel between MLS-B and pair approximation is at a conceptual level, not necessarily an exact mathematical one. (We shall see plenty of exact mathematics soon enough.) The essential fact is that both extend our thinking beyond the mean field, while both are themselves limited in the same way: context matters, but the relationships among contexts are neglected.

One can extend the basic notion of pair approximation to higher orders. Instead of reducing everything to pairs, we can describe an ecosystem using triples, for example. Mathematically, if we write a probability distribution over all possible states an ecosystem can be in, then imposing a mean-field approximation means replacing that with a product of many independent single-variable probability distributions. In other words, mean-field theory depends on the assumption that the joint probability distribution can be *factored* in a drastic way without losing important information. A pair approximation is the imposition of a slightly less drastic factorization, namely one into probabilities for pairs of variables, instead of single variables. Factorizations into higher-order probability distributions yield higher-order moment closures. (The *exact* analogue of the Simon *et al.* MLS-B model falls within this picture, but is not necessarily the pair approximation which we used to illustrate the basic concept.)

We note that the terms “type 1” and “type 2 multilevel selection” are also employed in the literature. However, their usage is not quite consistent [293], and since the

distinction which Smith *et al.* make is the most convenient for our purposes, we will employ the MLS-A and MLS-B designations for the categories they describe. (We will, in this way, bypass talk of “the level at which fitness is assigned” [292].) In this case, we are confronted with a tradeoff, and we choose attempting to reduce ambiguity over maintaining continuity with older writings.

9.2 Fisher's Fundamental Theorem

We have in this subject an interesting situation. The mathematics that has been employed to date has not been very elaborate. It involves a lot of algebra, some of it linear, much of it high-school. The challenge lies in the *stories told about the mathematics*, stories which have historically been muddled, overzealous in their claims, communicated at cross-purposes and, dare I say it, politicized. Later, we will try to sort some of that out. But before we reach that point, we will get the important equations in place.

Consider two populations, each composed of individuals of different types in various amounts. Let the proportion of type i in the first population be p_i , and denote the proportion of type i in the second population by p'_i . We can let the index i range from 1 to however large it needs to be to enumerate all types present in both populations. The basic statements of normalization are

$$\sum_i p_i = 1, \quad \sum_i p'_i = 1. \quad (9.1)$$

Now, introduce a third set of quantities to express the relationship between the two populations, thusly:

$$p_i \tilde{w}_i = p'_i. \quad (9.2)$$

This defines $\tilde{w}_i$ in terms of the given information about the population demographics. We suppose that this is always possible, *i.e.*, that there is no i for which $p_i = 0$ but $p'_i \neq 0$.

If we take the mean of $\tilde{w}_i$ across all types present in the first population, weighted by their abundances, we find that

$$\begin{aligned} \bar{\tilde{w}} &= \sum_i p_i \tilde{w}_i \\ &= \sum_i p'_i \\ &= 1. \end{aligned} \quad (9.3)$$

If we instead take the mean weighted by the proportions in the *second* population, we

9 The Varieties of Multilevel Selection

obtain

$$\begin{aligned}\bar{w}' &= \sum_i p'_i \tilde{w}_i \\ &= \sum_i p_i \tilde{w}_i^2.\end{aligned}\tag{9.4}$$

The difference between the two averages can be written

$$\bar{w}' - \bar{w} = \sum_i p_i \tilde{w}_i^2 - 1 = \sum_i p_i \tilde{w}_i^2 - \left(\sum_j p_j \tilde{w}_j \right)^2.\tag{9.5}$$

The right-hand side has the form of a variance, as we defined in Chapter 5: the second moment minus the square of the first. It is often called as such. And, if we designate $\tilde{w}_i$ the *relative fitness* of type i , then we can read Eq. (9.5) as saying that the change in fitness from one population to the other is the variance of fitness in the first population.

When we say it in words, this statement sounds like the definition of a dynamical system, or perhaps the beginnings of such a definition. However, it really isn't: we took both populations as given, and we declared that the $\{w_i\}$ have the right values to relate them. Eq. (9.5) holds by fiat. It is neither a predictive statement—given some $\{p_i\}$ and $\{p'_i\}$ obtained experimentally, it tells us nothing we did not already know—nor an update rule for a dynamical system that we can iterate and thereby explore mathematically.

The word “fitness” makes Eq. (9.5) sound like biology, and in that context, it is known as *Fisher’s fundamental theorem* [294]. But as an arithmetic identity, Eq. (9.5) has nothing necessarily to do with genetics or evolution. To illustrate, let the first population be a set of books, and let the second be a set of movie and TV adaptations. Many books never make the journey to the small or large screens. Some books are adapted once, and some are adapted multiple times. For example, Dashiell Hammett’s novel *The Maltese Falcon* (1929) was made into a movie by that name in 1931, another film titled *Satan Met a Lady* in 1936 and, most famously, the version starring Humphrey Bogart opposite Mary Astor in 1941. We could say that the novel’s fitness in Hollywood is $w_1 = 3$, if its index on our list is $i = 1$. Assigning “fitnesses” w_i in this way for all values of i , we can then define

$$\tilde{w}_i = \frac{w_i}{\sum_j p_j w_j},\tag{9.6}$$

and the sum of the $\tilde{w}_i$, weighted by p_i , will be unity. Fisher’s fundamental theorem tells us that the change in fitness due to moving from the written word to the moving picture is the fitness variance among the original novels.

This should make clear the difference between an arithmetic identity and a predictive model. The former can be “always true” or “universally applicable”; nevertheless, without additional assumptions, it doesn’t go anywhere. In turn, those additional assumptions can make the application of the identity invalid. For example, suppose

one wishes to compare two configurations of an ecosystem separated by a long interval of time. The ecosystem dynamics are defined in terms of short-term ecological or game-based interactions (as we did in Chapters 3 and 4). The fitnesses which arise as functions of game payoffs generally will not be the ones necessary to fill the role of the $\{w_i\}$ relating the two population configurations.

9.3 The Price Equation: Motivation and Shortcomings

Biological evolution is the change over generations of the genetic composition of populations due to natural factors, typically including significant randomness. Describing this mathematically, and developing quantitative tools to predict what might evolve under which conditions, is a great challenge. One place to begin is by describing, in a nice way, a population's change in genetic character from one generation to the next. By "a nice way", we mean that we'd like to be able to attribute changes to the appropriate influences. What changes are due to random mutations creating new variations, for example, and what changes are due to natural selection winnowing out varieties which cannot survive in their environment?

We can make a crude measure of a population's genetic composition by counting up how many organisms in the population have a certain gene of interest. We can express this amount as a percentage of the total population, saying, for example, "The frequency of gene A in this population is 0.22." This, of course, is a mean-field statement. We know that such statements can be insufficient for making viable predictions about *dynamics*, but in this section, we will assume a more modest aim, and try only to manipulate the *description*, as we did in deriving Fisher's theorem.

In this section, we use the notation of van Veelen [295].

We consider two populations, S_1 and S_2 . All the offspring of organisms in S_1 belong to S_2 , and all the parents of organisms in S_2 are in S_1 . We write N for the size of population S_1 . For an individual $i \in S_1$, the frequency of gene A is

$$q_i = \frac{g_i}{l_z}, \quad (9.7)$$

where l_z is the zygotic ploidy, *i.e.*, the number of copies of each chromosome carried by a fertilized egg. The frequency of gene A in population S_1 is

$$Q_1 = \frac{\sum_{i \in S_1} g_i}{l_z N} = \frac{\sum_i q_i}{N}. \quad (9.8)$$

We want to relate Q_2 and Q_1 . One simple way to do so is to take their difference:

$$\Delta Q = Q_2 - Q_1. \quad (9.9)$$

We can write Q_2 as

$$Q_2 = \frac{\sum_i g'_i}{l_g \sum_i z_i}, \quad (9.10)$$

9 The Varieties of Multilevel Selection

where l_g is the gametic ploidy, z_i is the number of successful gametes from individual i , and g'_i is the number of A -type genes in the set of all successful gametes from individual i . The proportion of A -type genes in that set is

$$q'_i = \frac{g'_i}{z_i l_g}. \quad (9.11)$$

From this,

$$Q_2 = \frac{\sum_i z_i l_g q'_i}{l_g \sum_i z_i} = \frac{\sum_i z_i q'_i}{\sum_i z_i}. \quad (9.12)$$

Therefore,

$$\Delta Q = \frac{\sum_i z_i q'_i}{\sum_i z_i} - \frac{1}{N} \sum_i q_i. \quad (9.13)$$

We'd like our expression for the change in Q to be written in terms of the changes in the individual q_i , so we subtract and add a sum over q_i :

$$\Delta Q = \frac{\sum_i z_i (q'_i - q_i)}{\sum_i z_i} + \frac{\sum_i z_i q_i}{\sum_i z_i} - \frac{1}{N} \sum_i q_i. \quad (9.14)$$

Next, we gather the last two terms over a common denominator:

$$\Delta Q = \frac{\sum_i z_i (q'_i - q_i)}{\sum_i z_i} + \frac{\sum_i z_i q_i - \frac{1}{N} \sum_i q_i \sum_j z_j}{\sum_i z_i}. \quad (9.15)$$

Now, we factor an N out of the latter term.

$$\Delta Q = \frac{\sum_i z_i (q'_i - q_i)}{\sum_i z_i} + \frac{N}{\sum_i z_i} \left[\frac{1}{N} \sum_i z_i q_i - \frac{1}{N^2} \sum_i q_i \sum_j z_j \right]. \quad (9.16)$$

We rearrange this just a bit to yield the **Price equation**:

$$\Delta Q = \frac{N}{\sum_i z_i} \left[\frac{1}{N} \sum_i z_i q_i - \left(\frac{1}{N} \sum_i q_i \right) \left(\frac{1}{N} \sum_j z_j \right) \right] + \frac{\sum_i z_i (q'_i - q_i)}{\sum_i z_i}. \quad (9.17)$$

This is just an algebraic identity: we took the compositions of the two populations as given, and we wrote a fancy expression for the change of gene frequency between them. *We have not said anything about dynamics from which this change could be derived, nor have we made any claims about what changes are more probable than others.* Eq. (9.17) is a rearrangement of the given information, not an update rule for a dynamical system. It is not even a statement about probabilities, although the expression in brackets formally resembles the covariance between two random variables. Nothing in the derivation of Eq. (9.17) was a random variable; despite this, the Price

9.3 The Price Equation: Motivation and Shortcomings

equation is typically written in “covariance” notation. This poor tradition of notation has contributed more than a little confusion to the subject.

Van Veelen *et al.* [296] make the point in the following way:

[W]hat is most important is that we realize that the numerical input of the Price equation is a list of numbers. It is a list that concerns two generations, and which tracks who is whose offspring. But whatever it reflects, it is crucial to realize that the point of departure is nothing but a list of numbers. This list of numbers is used twice. First we use it to compute the frequencies of the gene under consideration in generations 1 and 2, respectively, and subtract the latter from the former. This amounts to the change in gene frequency. Then we use the same list to compute a few other, slightly more complex quantities. The essence of the Price equation is that these quantities also add up to the change in gene frequency. One way of computing the change in frequency therefore can be rewritten as the other and vice versa. What they are, therefore, is nothing but two equivalent ways to compute the change in gene frequency, given a list of numbers concerning genes in two subsequent generations ... Whether this particular second generation is likely to follow the first or not, the two ways of computing the change in frequency return the same number.

To make a physics analogy, what we have done is like starting with Newton’s second law, $\vec{F} = m\vec{a}$, and writing it as

$$m\vec{a} = m\vec{a}. \quad (9.18)$$

We could then rewrite the $\vec{a}$ vectors in some elaborate way. For example, we could write one side of the equation in Cartesian coordinates and the other in spherical coordinates, giving some complicated formulas involving trigonometric functions all over the place. These formulas would be *true*, in the sense that Euclidean geometry is *true*, but they would contain no *physics*. In some circumstances, they might be useful, but we could not wring value out of them without some extra assumptions about the dynamics at work.

We now make a series of assumptions geared towards turning the Price equation (9.17) into something more like an update rule for a dynamical system.

First, we specialize our considerations to a scenario in which individuals interact by donating effort or assistance to one another. Donors of effort increase the number of successful gametes produced by the recipient, at the expense of their own. We parameterize this in the following way: denote by c a donor’s decrease in successful gametes of its own, and denote by b the increase in successful gametes of the recipient. We idealize interactions as pairwise events, and so we keep track of them using matrices. The first index, i , denotes an individual in population S_1 . The second index, α , ranges over the occasions on which interactions can take place.

$$\Delta Q = \frac{N}{\sum_i z_i} \left[\frac{1}{N} \sum_i z_i q_i - \left(\frac{1}{N} \sum_i q_i \right) \left(\frac{1}{N} \sum_j z_j \right) \right] + \frac{\sum_i z_i (q'_i - q_i)}{\sum_i z_i}. \quad (9.19)$$

9 The Varieties of Multilevel Selection

We can be slightly more general and allow each individual to have their own ploidy, l_i . So, instead of using the population size N , we use $\sum_i l_i$. Following the literature, we calculate the number of successful gametes per haploid set, $w_i = z_i/l_i$.

$$\Delta Q = \frac{\sum_i l_i}{\sum_i l_i w_i} \left[\frac{\sum_i l_i w_i q_i}{\sum_i l_i} - \left(\frac{\sum_i l_i w_i}{\sum_i l_i} \right) \left(\frac{\sum_i l_i q_i}{\sum_i l_i} \right) \right] + \frac{\sum_i l_i w_i (q'_i - q_i)}{\sum_i l_i w_i}. \quad (9.20)$$

We now make two additional assumptions:

1. The second term in this form of the Price identity is negligible.
2. The fitnesses z_i can be written

$$z_i = l_i w_i = f_i + b \sum_{\alpha} S_{i\alpha} - c \sum_{\alpha} Q_{i\alpha}. \quad (9.21)$$

Here, $\sum_{\alpha} S_{i\alpha}$ is the total number of times individual i received a benefit, and $\sum_{\alpha} Q_{i\alpha}$ is the number of times individual i incurred a cost.

We introduce the abbreviation

$$\bar{q} = \frac{\sum_i l_i q_i}{\sum_i l_i}. \quad (9.22)$$

Dropping the last term of ΔQ and substituting in our chosen form for $l_i w_i$, we arrive after some algebra at the following:

$$\Delta Q = \left(\frac{\sum_{i,\alpha} Q_{i\alpha} (q_i - \bar{q})}{\sum_i l_i w_i} \right) \left[\left(\frac{\sum_{i,\alpha} S_{i\alpha} (q_i - \bar{q})}{\sum_{i,\alpha} Q_{i\alpha} (q_i - \bar{q})} \right) b - c \right]. \quad (9.23)$$

The quantity in square brackets has the form of Hamilton's condition, if we identify the quotient multiplying b as a measure of assortment:

$$r = \frac{\sum_{i,\alpha} S_{i\alpha} (q_i - \bar{q})}{\sum_{i,\alpha} Q_{i\alpha} (q_i - \bar{q})}. \quad (9.24)$$

9.4 Interconverting MLS-A and Inclusive Fitness

Hamilton [297] defines inclusive fitness in the following manner.

Inclusive fitness may be imagined as the personal fitness which an individual actually expresses in its production of adult offspring as it becomes after it has been first stripped and then augmented in a certain way. It is stripped of all components which can be considered as due to the individual's social environment, leaving the fitness which he would express if not exposed to any of the harms or benefits of that environment. This quantity is then augmented by certain fractions of the quantities of harm and benefit which the individual himself causes to the fitnesses of his neighbours. The

fractions in question are simply the coefficients of relationship appropriate to the neighbours whom he affects: unity for clonal individuals, one-half for sibs, one-quarter for half-sibs, one-eighth for cousins, [...] and finally zero for all neighbours whose relationship can be considered negligibly small.

Allen [151] comments,

At this point you may be asking, “Wait, does it really make sense to divide offspring into those produced on one’s own versus those produced by help from others?” This is exactly the problem! Aside from the obvious point that no one reproduces without help in sexual species, nature is full of synergistic and nonlinear interactions, so that making clean divisions like this is impossible in most situations. Thus the idea of inclusive fitness theory only works in simplified toy models of reality.

Inclusive fitness has variously been claimed to supersede multilevel selection, to be mathematically equivalent to MLS or to be a subset of it. In order to clarify the relationship between MLS and inclusive fitness (IF) models, it helps to have a specific example in hand. Fortunately, Bijma and Wade [46] have provided an explication which, made slightly more general, is quite helpful.

The remainder of this section will be specific to MLS-A. That is, while we will sometimes speak of “groups” within the population, these groups will be formed at random from the pool of available individuals, rather than being entities which have their own explicit dynamics included in the update rules. Furthermore, we will consider only short-term changes, comparing one generation to the next, instead of entire evolutionary trajectories. This is just the kind of comparison where we can apply the Price equation which we derived in the previous section.

We shall consider a quantitative genetic model in which the trait value of an individual organism affects its fitness as well as the fitness of those with which it interacts. The trait value of individual i , which we denote P_i , depends on a genetic component and a nonheritable, environmental component:

$$P_i = G_i + E_i. \quad (9.25)$$

The personal fitness of individual i depends on that individual’s trait value, P_i , as well as the trait values of those in its social group. For simplicity, we imagine that all groups have size 2; that is, each individual i interacts with a partner j . In addition, we assume that the effect of trait value on fitness is linear, and that the combination of the self-value and partner-value effects is also linearly additive. Including residual effects (due, say, to exogenous environmental variation), we can write

$$W_i = \text{const.} + \beta_{i,i}P_i + \beta_{i,j}P_j + e_i. \quad (9.26)$$

It could be that, for the evolutionary ecosystem one is trying to model, the relationship between trait value and fitness is not linear. One could still run a linear regression on experimental data obtained from that system, but the output from the statistical

9 The Varieties of Multilevel Selection

software package would not be meaningful [47]. We shall neglect this complication for the moment and return to it in a later section.

The next step in model-building would be to define a rule which gives the following generation in terms of the current one and the set of fitness values $\{W_i\}$. Typically, one uses the Price equation to do so, saying that “the change in gene frequency is given by the covariance of genes and fitness.” This amounts to pretending that a purely descriptive equation has predictive value. In other words, taking this step requires importing additional assumptions. (The language of the invocation, with its talk of “covariance,” also propagates the confusion about the Price equation and its status, which we touched upon in the previous section.)

Furthermore, the update rule one invents by this scheme does not even really define a dynamical system: it requires more information about the current generation than it can yield about the next generation, so it cannot be iterated. We can say that this update rule is not “dynamically sufficient.”

This “inclusive” or “neighbor-modulated” fitness calculation, Eq. (9.26), represents the state of the population by the trait values $\{P_i\}$. We can equally well use a group selection calculation—that is, an MLS-A model—wherein we say that the personal fitness of individual i depends on the mean trait value of its social circle,

$$\bar{P}_g = \frac{1}{2}(P_i + P_j), \quad (9.27)$$

and on how far P_i deviates from that average,

$$\Delta P_i = P_i - \bar{P}_g = \frac{1}{2}(P_i - P_j). \quad (9.28)$$

If we once again assume linear relationships, we can write the personal fitness W_i in terms of a between-group component and a within-group component:

$$W_i = \text{const.} + e_i + \beta'_{i,g}\bar{P}_g + \beta'_{i,\Delta}\Delta P_i. \quad (9.29)$$

If we substitute our definitions of $\bar{P}_g$ and ΔP_i into this equation, we find

$$\begin{aligned} W_i &= \text{const.} + e_i + \beta'_{i,g}\frac{P_i + P_j}{2} + \beta'_{i,\Delta}\frac{P_i - P_j}{2} \\ &= \text{const.} + e_i + \frac{1}{2}(\beta'_{i,g} + \beta'_{i,\Delta})P_i + \frac{1}{2}(\beta'_{i,g} - \beta'_{i,\Delta})P_j. \end{aligned} \quad (9.30)$$

Comparing this expression with the “inclusive” or “neighbor-modulated” fitness expression in Eq. (9.26), we see that

$$\beta_{i,i} = \frac{1}{2}(\beta'_{i,g} + \beta'_{i,\Delta}), \quad (9.31)$$

and

$$\beta_{i,j} = \frac{1}{2}(\beta'_{i,g} - \beta'_{i,\Delta}). \quad (9.32)$$

9.4 Interconverting MLS-A and Inclusive Fitness

On the one hand, we have the formula based on inclusive fitness, Eq. (9.26), and on the other, we have the levels-of-selection formula, Eq. (9.29). The relation between the two is a *linear transformation of coordinates*: we change from one way of tallying trait values to another. When we transform our coordinate system, the parameters of our model get mixed up with one another, as seen in Eqs. (9.31) and (9.32).

A mechanical analogy is illustrative: if we wish to study the collision of two billiard balls on a table, we can look at the collision in countless different ways. We could describe what happens from the perspective of an observer standing at rest with respect to the pool table. Or, alternatively, we could view the situation from the perspective of an observer who is at rest with respect to the center of mass of the two billiard balls. We can even switch perspectives in the middle of a calculation: starting with information given in terms of the table rest frame, we transform into the center-of-mass rest frame to see what Newton's laws imply, and then we transform back into the table rest frame to predict what the observer standing beside the table will see. The biological situation is analogous: we have transformed the trait values from a laboratory reference frame into a “center of group” frame.

We can express this transformation more generally using matrix notation. If we write the trait values as a vector, then the group-based trait values are given by the vector of individual trait values multiplied by a matrix:

$$\begin{pmatrix} \bar{P}_g \\ \Delta P_i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} P_i \\ P_j \end{pmatrix}. \quad (9.33)$$

Writing $\mathbf{\Lambda}$ for the transformation matrix,

$$\mathbf{\Lambda} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (9.34)$$

we are saying that the MLS-A model's trait values, $\mathbf{P}'$, are related to the inclusive-fitness trait values, $\mathbf{P}$, by the equation

$$\mathbf{P}' = \mathbf{\Lambda P}. \quad (9.35)$$

We know that the results of the two calculations must agree. In matrix form, this requirement means

$$\beta' \mathbf{P}' = \beta \mathbf{P}, \quad (9.36)$$

where the matrix β is

$$\beta = \begin{pmatrix} \beta_{i,i} & \beta_{i,j} \\ \beta_{j,i} & \beta_{j,j} \end{pmatrix}. \quad (9.37)$$

By substituting in the transformation rule for $\mathbf{P}$, we see that

$$\beta'(\mathbf{\Lambda P}) = \beta \mathbf{P}, \quad (9.38)$$

which in turn means that

$$\beta = \beta' \mathbf{\Lambda}. \quad (9.39)$$

9 The Varieties of Multilevel Selection

This is the matrix version of Eqs. (9.31) and (9.32).

The matrix-algebra statement of the relationship between IF and MLS-A, although somewhat more abstract than the original equations, brings forth the essential point: when we change the way we describe trait values, we must make a corresponding change in the parameters which control the evolutionary dynamics. If we go from $\mathbf{P}$ to $\mathbf{P}'$ by the transformation $\mathbf{\Lambda}$, then we change our coefficients β by the inverse of $\mathbf{\Lambda}$:

$$\beta' = \beta \mathbf{\Lambda}^{-1}. \quad (9.40)$$

Because our IF model and our MLS-A model are different perspectives on the same mathematics, problems with one will be problems with the other, only seen from a different angle. For example, to predict what strategy will be evolutionarily favored, the IF model relies on Hamilton's concept of "relatedness," and in practical scenarios, "relatedness" is problematic. But if "relatedness" is ill-defined or inapplicable, mixing it with another parameter by means of a coordinate transformation will do no good, either.

9.5 An Example of MLS-B

In order to understand how MLS-B goes beyond MLS-A, we now turn to an example of an MLS-B dynamical system, constructed by van Veelen *et al.* [290]. Per the definition of MLS-B, the population is organized into groups, and the dynamical update rules make explicit reference to the group level of structure. In this particular model, all groups are taken to have the same size, and this size is constant over time. This is accomplished by balancing reproduction events with death events. If an individual within a group reproduces, an organism is picked at random from that group and killed. Organisms come in two types, which we designate *cooperators* and *defectors*. The difference between the types manifests in a difference in reproductive fecundities. For simplicity, we say that cooperators reproduce at rate 1, while defectors reproduce at an augmented rate, $1 + s$.

In order to make this an MLS-B model, we need an explicit group-level dynamic. So, we say that in addition to the within-group reproduction of individuals, *entire groups also reproduce*. The offspring group has the same proportion of cooperators and defectors as its parent group. The rate of group reproduction depends on that proportion: if the size of the group is k and it contains n_c cooperators, then that group reproduces at a rate $1 + u \left(\frac{n_c}{k} \right)^\alpha$. The parameter α controls the degree of nonlinearity in the dynamics.

As van Veelen *et al.* observe,

Being a cooperator therefore comes at a cost—it reduces the reproduction rate of the individual by s —but it has a benefit for all group members, including itself, through an increase in the rate at which the group as a whole reproduces.

Taking the limit of large population size and large k , the dynamics of this system can be cast into a deterministic partial differential equation. Let x be the fraction of

cooperators in a group, and denote by $\mu(x; t)$ the relative frequency of groups having cooperator fraction x at time t . Then,

$$\frac{d\mu(x; t)}{dt} = s \frac{d}{dx} [x(1-x)\mu(x; t)] + u\mu(x; t) \left[x^\alpha - \int_0^1 dy y^\alpha \mu(y; t) \right]. \quad (9.41)$$

The right-hand side of this PDE has two terms, reflecting two effects at work. The first term, with its logistic form, describes the decrease of x within a group, due to the defectors' local advantage. The second term, proportional to the parameter u , indicates how groups with low x decrease in frequency, and those with high x increase in frequency, due to natural selection. The two contributions of opposite sign within the brackets ensure that the overall population size remains constant.

Treating $\mu(x; t)$ as a time-dependent probability density function, we can summarize the population by the moments of x :

$$\langle x^n(t) \rangle = \int_0^1 dx x^n \mu(x; t). \quad (9.42)$$

The overall frequency of cooperators changes at a rate that can be computed, after some algebra:

$$\frac{d\langle x \rangle}{dt} = s(\langle x^2 \rangle - \langle x \rangle) + u(\langle x^{\alpha+1} \rangle - \langle x \rangle \langle x^\alpha \rangle). \quad (9.43)$$

The simplest case is where $\alpha = 1$, and the cooperators' effect on group reproduction rate is linear. Then,

$$\frac{d\langle x \rangle}{dt} = s(\langle x^2 \rangle - \langle x \rangle) + u(\langle x^2 \rangle - \langle x \rangle^2). \quad (9.44)$$

The s term, which reflects within-group selection, will be negative or zero. The u term is zero or positive, and it indicates the effect of *between-group* selection. Whether the frequency of cooperation increases or decreases depends on which of these two countervailing influences wins out.

Rewriting the previous equation, we find that

$$\frac{d\langle x \rangle}{dt} = u(\langle x^2 \rangle - \langle x \rangle^2) + s(\langle x^2 \rangle - \langle x \rangle) \quad (9.45)$$

$$= (\langle x^2 \rangle - \langle x \rangle^2)(s + u) - (\langle x \rangle - \langle x \rangle^2)s \quad (9.46)$$

$$= \langle x \rangle (1 - \langle x \rangle) \left[\left(\frac{\langle x^2 \rangle - \langle x \rangle^2}{\langle x \rangle - \langle x \rangle^2} \right) (s + u) - s \right]. \quad (9.47)$$

This is reminiscent of when we turned the Price equation into a condition resembling Hamilton's rule, in Eqs. (9.23) and (9.24).

Note that the parameter s is the "cost of cooperation," in that it indicates the demotion of the reproductive rate due to being a cooperator rather than a defector. Likewise, the aggregate benefits that other group members accrue thanks to the pres-

9 The Varieties of Multilevel Selection

ence of a cooperator are $b = s + u$. The s is from the reduction in the death rate, due to the balancing of birth and death events in each group. So, if we write the assortment coefficient r as

$$r = \frac{\langle x^2 \rangle - \langle x \rangle^2}{\langle x \rangle - \langle x \rangle^2}, \quad (9.48)$$

then we can write the rate of change of $\langle x \rangle$ as

$$\frac{d\langle x \rangle}{dt} = \langle x \rangle (1 - \langle x \rangle)(rb - c). \quad (9.49)$$

The sign of $d\langle x \rangle/dt$ depends on a factor that has the form of Hamilton’s rule. Remember, though, that in order to achieve this form, we had to define r in terms of the current population demographics, by way of the moments $\langle x \rangle$ and $\langle x^2 \rangle$. This means that, as the population composition changes, r will vary. We could try to apply Hamilton’s rule to solve the problem, but in order to apply Hamilton’s rule, we have to solve the problem anyway!

When $\alpha = 1$, we see that we can express the change in cooperator frequency by an equation with a natural MLS interpretation, or by a formula with an IF flavor. The two are freely interconvertible. *However*, if we introduce nonlinearity by choosing $\alpha > 1$, we still have the basic MLS expression (9.43), but we can no longer reshape it into an IF condition. Should we try to define an assortment coefficient r as we did for the $\alpha = 1$ case, we would find that it necessarily depends on all the parameters of the dynamics. That is, our r would include s , u and α , in addition to the moments of x .

There’s no reason why MLS-B should be the end of the story. Earlier, we noted that MLS-B is conceptually analogous to a pair approximation. We have already seen multiple reasons why pair approximations, and similar but more elaborate extensions beyond mean-field theory, fail to capture important effects. This was a recurring motif in Chapter 3, where competition was defined in terms of predation on a limited resource. We also saw it in Chapter 4, when the effects of network topology manifested themselves.

9.6 A Literature of Confusion

In §8.4, we quoted Damore and Gore’s point that errors propagate through the works in this area, thanks to assumptions going unspoken and definitions being swept into the Supplemental Information. The evolutionary theory of social behaviors is a field where there is no substitute for reading the primary sources.

For example, Dawkins recently claimed that Hölldobler has “no truck with group selection” [298]. A 2005 piece by Wilson and Hölldobler proposes, in the first sentence of its abstract, that “group selection is the strong binding force in eusocial evolution” [299]. Later, Hölldobler (with Reeve) voiced support for the “trait-group selection and individual selection/inclusive fitness models are interconvertible” attitude [149]. Hölldobler’s book with Wilson, *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* [300], maintains this tone. Quoting from page 35:

It is important to keep in mind that mathematical gene-selectionist (inclusive fitness) models can be translated into multilevel selection models and vice versa. As Lee Dugatkin, Kern Reeve, and several others have demonstrated, the underlying mathematics is exactly the same; it merely takes the same cake and cuts it at different angles. Personal and kin components are distinguished in inclusive fitness theory; within-group and between-group components are distinguished in group selection theory. One can travel back and forth between these theories with the point of entry chosen according to the problem being addressed.

This is itself a curtailed perspective, whose validity is restricted to a narrow class of implementations of the “multilevel selection” idea. Regardless, this is not at all equivalent with having “no truck with group selection”. The statement “method *A* is no better or worse than method *B*” is a far cry from “method *A* is worthless and only method *B* is genuinely scientific”.

We have also a 2010 solo-author piece by Hölldobler, in a perspective printed in *Social Behaviour: Genes, Ecology and Evolution* [301]. Quoting from page 127:

I was, and continue to be, intrigued by the universal observation that wherever social life in groups evolved on this planet, we encounter (with only a few exceptions) a striking correlation: the more tightly organized within-group cooperation and cohesion, the stronger the between-group discrimination and hostility. Ants, again, are excellent model systems for studying the transition from primitive eusocial systems, characterized by considerable within-group reproductive competition and conflict, and poorly developed reciprocal communication and cooperation, and little or no between-group competition, on one side, to the ultimate superorganisms (such as the gigantic colonies of the *Atta* leafcutter ants) with little or no within-group conflict, pronounced caste systems, elaborate division of labour, complex reciprocal communication, and intense between-group competition, on the other side (Hölldobler & Wilson 2008 [the book quoted above]).

And, a little while later, on p. 130:

In such advanced eusocial organisations the colony effectively becomes a main target of selection [...] Selection therefore optimises caste demography, patterns of division of labour and communication systems at the colony level. For example, colonies that employ the most effective recruitment system to retrieve food, or that exhibit the most powerful colony defence against enemies and predators, will be able to raise the largest number of reproductive females and males each year and thus will have the greatest fitness within the population of colonies.

Hölldobler also says that these phenomena can be thought of as “extended phenotypes,” which is a Dawkinsian turn of phrase; this is consistent with the “MLS and IF are interchangeable” theme.

One author who has been less often read and more often misread third-hand is the zoologist V. C. Wynne-Edwards [302, 303]. In the early 1960s, Wynne-Edwards compiled a thick volume documenting how different species use signaling mechanisms to determine their population density and, effectively, figure out how crowded the living situation is. Introducing the compendium, he said that group selection would have to be the explanation for these observations, since, in essence, that's how cooperative behaviors are explained. To quote an essay [304] that he wrote shortly after,

In a recent book [302] I advanced a general proposition which may be summarized in the following way. (1) Animals, especially in the higher phyla, are variously adapted to control their own population densities. (2) The mechanisms involved work homeostatically, adjusting the population density in relation to fluctuating levels of resources; where the limiting resource is food, as it most frequently is, the homeostatic system prevents the population from increasing to densities that would cause over-exploitation and the depletion of future yields. (3) The mechanisms depend in part on the substitution of conventional prizes, namely, the possession of territories, homes, living space and similar real property, or of social status as the proximate objects of competition among the members of the group concerned, in place of the actual food itself. (4) Any group of individuals engaged together in such conventional competition automatically constitutes a society, all social behaviour having sprung originally from this source.

Wynne-Edwards' point (2) is reminiscent of the descendant-shading effect we saw in Chapter 3. Indeed, one can add social signalling among consumers to the host-consumer model, and it turns out that curtailing consumption in overcrowded situations is a robustly evolvable trait [72]. However, in the spatial host-consumer model, Wynne-Edwards' point (3) is not necessary: the consumers do not substitute any "social status" or "conventional prize" in the place of the actual resource, but reproductive restraint evolves anyway.

Wynne-Edwards continues,

In developing the theme it soon became apparent that the greatest benefits of sociality arise from its capacity to override the advantage of the individual members of in the interests of the survival of the group as a whole. The kind of adaptations which make this possible, as explained more fully here, belong to and characterize social groups as entities, rather than their members individually. This in turn seems to entail that natural selection has occurred between social groups as evolutionary units in their own right, favouring the more efficient variants among social systems wherever they have appeared, and furthering their progressive development and adaptation.

The general concept of intergroup selection is not new. It has been widely accepted in the field of evolutionary genetics, largely as a result of the classical analysis of Sewall Wright [43, 305, 306]. He has expressed the view that "selection between the genetic systems of local populations of a

species ... has been perhaps the greatest creative factor of all in making possible selection of genetic systems as wholes in place of mere selection according to the net effect of alleles" [43]. Intergroup selection has been invoked also to explain the special case of colonial evolution in the social insects [307, 308, 309].

Others criticized Wynne-Edwards using "group selection" models which we can classify as MLS-A. These critiques, however, impose upon Wynne-Edwards a conception of population structure that was not his own [303]. Indeed, the image of population structure that Wynne-Edwards appears to have in mind is closer to that realized in a cellular automaton lattice model, than to that seen in the critiques by Maynard Smith [310, 311] and others.

We conclude this section by revisiting Fisher's fundamental theorem, which we developed as an arithmetic identity in §9.2. Fisher himself [176] stated what he called "the fundamental theorem of Natural Selection" in the following way:

The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.

Fisher arrives at this statement after a lengthy discussion which invokes many more particular assumptions than we did in §9.2. Moreover, he takes his fundamental theorem to be an approximation:

Since the theorem is exact only for idealized populations, in which fortuitous fluctuations in genetic composition have been excluded, it is important to obtain an estimate of the magnitude of the effect of these fluctuations, or in other words to obtain a standard error appropriate to the calculated, or expected, rate of increase in fitness.

Fisher then considers a model of a panmictic population and argues using that model that fluctuations should decrease with the square root of the population size.

In other words, although Fisher was quite favorably impressed with the importance of his result, he himself did not take it as universally true. This is in sharp contrast to more recent authors who take the Price equation as *the* way to think about evolutionary change, thereby imparting a glow of perceived universality to Fisher's theorem.

9.7 Discussion

Consider the following statement from a recent popular-science summary of the perennial MLS/IF dispute [312]:

In the final analysis, multilevel selection is little more than a rebranding of Hamilton's inclusive fitness (albeit the "enhanced" 1975 version).

That such a claim can be the punchline of a popularization is definitely an advance from the curmudgeonly attitude that only IF is viable and anything which sounds like MLS is old-fashioned at best or antiscientific at worst. However, it is only the first step

9 The Varieties of Multilevel Selection

in a rhetorical journey. We have seen that this interchangeability is straightforwardly true in mean-field theory, the domain of MLS-A. Nevertheless, there are more things in heaven and earth than are dreamt of in that approximation, and it is well past time to move on to new adventures.

If one defines MLS so narrowly that it *is* indeed just IF in a new coordinate system, then MLS will inherit all the problems which limit IF's usefulness in modern biology. On the other hand, if one defines MLS broadly, then one invites misunderstanding from those who, knowingly or not, define it narrowly. Likewise, if we try to reserve the term "group selection" for those cases which include explicit group-level processes (MLS-B but *not* MLS-A), then we engender the same confusion.

The esteemed science communicator Larry Gonick [313] identified the key problem:

We always overestimate the degree to which we are understood! That is, when I talk, I assume you understand me—unless you tell me otherwise! And there are a lot of reasons you may not give me this essential feedback: you're too polite; you thought you got it, even though you didn't; you're afraid of looking stupid. The result is a worldwide overvaluation of the level of understanding!

In this chapter, we have used the designations MLS-A and MLS-B. A future development might conceivably push the terminology into the alphanumeric, perhaps introducing "MLS-114" and the like. Jargon of this kind is intimidating and appears quite nontransparent. It is the sort of insider-speak which fills a room with fog. But this fog comes with a silver lining: we may not know what "MLS-B" means when we hear it, but we *know that we do not know*. This is unlike what happens when we hear a nontechnical-sounding term like "group selection"—the language may seem more friendly, but that sense of welcome is dangerous. Its seeming hospitality leads to miscommunication and ceaseless confusion. Technical argot befuddles the outsider and, all too often, the student, but thoughtfully chosen terminology can do wonders to make one's meaning clear, at least to fellow professionals. (The words which doctors throw around are intimidating, particularly when we are patients ourselves, but they do have good reasons to use the language they learned in med school: drawing careful distinctions which everyday speech cannot save lives.) Perhaps we must accept the cost of jargon and mathematics, until such time as we have cleared enough of that confusion that we can make the essential points in plainer speech without error. If our choice is between an algebra prerequisite and another generation wasted on Team Inclusive Fitness versus Team Multilevel Selection, we unhesitatingly affirm our preference for the former.

10 Speculations for New Mathematics

According to John Baez [314],

The raw beginner in mathematics wants to see the solutions of an equation.
The more advanced student is content to prove that the solution exists. But
the master is content to prove that the equation exists.

This chapter is devoted to demonstrating that equations could, potentially, be written.

10.1 More on Multiscale Information

Our first major topic, the information-theoretic formalism for multiscale structure, presents us with several possibilities for future extensions. We saw that the complexity profile $C(k)$ depends on the amounts of information associated with all the possible subsets of a system. If $Q(j)$ denotes the sum of the joint information of all collections having size j ,

$$Q(j) = \sum_{i_1, \dots, i_j} H(a_{i_1}, \dots, a_{i_j}), \quad (10.1)$$

then the complexity at scale k is

$$C(k) = \sum_{j=N-k}^{N-1} (-1)^{j+k-N} \binom{j}{j+k-N} Q(j+1). \quad (10.2)$$

We saw that we can simplify this expression in the special case where all sets of the same size carry the same amount of information. This is a strong symmetry requirement, and one naturally wonders whether other, less stringent symmetries can also lead to useful and mathematically interesting expressions for $C(k)$.

Moreover, the Marginal Utility of Information has a conceptual connection with a method for detecting *community structure* within networks [315], and this relationship could potentially be developed further. An intuitive definition of a *module* within a network could be phrased in the following way: a set of nodes more strongly connected among themselves than they are to anything else [316]. The extent to which a network is “modular” would be, then, the extent to which it consists of multiple more-or-less independent pieces. We could try to formalize this intuition with some quality index for a division of a network into putative modules. If a partition of the vertices breaks only a few edges, then the quality of that partition is high. We could become more sophisticated and compare the number of severed connections with some null hypothesis, and then accord higher quality rankings to those partitions which cut

fewer edges than we'd expect by random chance. For example, if we have an undirected graph with E edges and we know the degrees k_i and k_j of two nodes, then our best guess for the number of edges linking the vertices i and j is $\frac{k_i k_j}{2E}$. The quality index Q would then be a comparison of this quantity to A_{ij} , the actual number of edges; we could then embark on a scheme to optimize Q . Such optimization problems turn out to be **NP**-hard [316], and while various heuristics appear to work well enough in networks one sees in practice, we should take care before calling their results the "best" partition, as they may be only first among equals in partition space [317].

An alternative, complementary perspective on modularity takes a more dynamical view. Suppose that each vertex in a graph comes with an oscillator attached (perhaps a pendulum or a firefly), and that oscillators for linked vertices are coupled. If the dynamics tend to synchrony, we'd expect that oscillators within a module would synchronize with each other first; the timescale for different modules to begin oscillating in phase would be longer. Or, we could unleash drunkards upon the network: a random walk starting within a module would be more likely to stay within that module than to escape and begin perambulating elsewhere. Turning this intuition around, we could *define* a module as a subset of nodes which a random walker is more apt to stay within than to leave. We could then take this approach and devise a procedure for partitioning a graph based on the conditional probability distribution

$$G_{ij}^t \equiv p(j|i, t), \quad (10.3)$$

denoting the probability that a random walker starting at i will be standing at j after a time period t . (The time period is typically taken to be the inverse of the smallest non-zero Laplacian eigenvalue, but there are other ways to find a good timescale that might be more computationally efficient.)

Because of the **NP**-hardness issue mentioned earlier, we should not expect to be able to write down a dynamical rule which, when implemented on an arbitrary graph, will yield a steady-state configuration revealing the optimal partition. Instead, we ought to expect metastable configurations, long decay timescales and fuzzy assignments of nodes to clusters.

Ziv, Middendorf and Wiggins (ZMW) use the diffusion of random walkers to calculate a conditional probability distribution $p(y|x)$. Both Y and X range over the vertices of the network we're studying; y gives the position which a random walk started at x reaches after a specified time. The description variable Z is an index over modules. The cardinality $|Z|$ indicates the number of modules we are considering in our network partitioning, and $p(z|x)$ is the probability that vertex x is mapped into module z . With these definitions, ZMW use the information bottleneck procedure to find optimal network partitions, *i.e.*, divisions into modules which best preserve $p(x, y)$.

One result of the ZMW procedure is a curve of information gained versus information provided, that is, of how well a description with a certain information content reproduces the random-walk probability distribution of the network. This curve is, essentially, the utility of a description as a function of description size. Each partition solution $p(z|x)$ has a Shannon information content $I(Z; X)$, which for comparison pur-

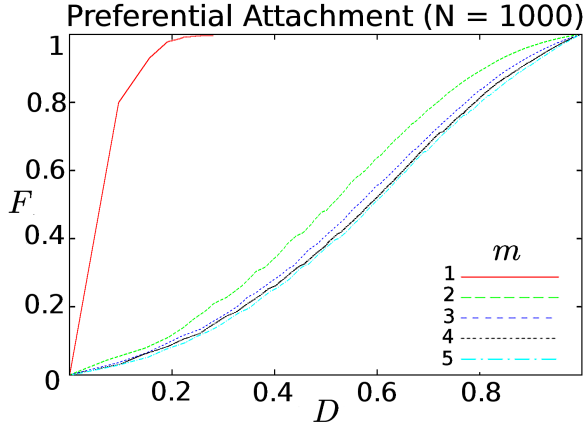


Figure 10.1: Fidelity curves for networks built by preferential attachment. The derivatives of these curves are analogous to the MUI.

poses we can normalize by $H(X)$. The fidelity of this module assignment scheme is the ratio of how well it predicts diffusion, $I(Z; Y)$, to how the original network did so, $I(X; Y)$. The *information curve* is a plot of $I(Z; Y)/I(X; Y)$ against $I(Z; X)/H(X)$. The ZMW paper presents such curves for some synthetic and empirical graphs, but knowing how the fidelity curves behave for typical classes of synthetic networks would provide much helpful context.

The first unexpected thing happened when I looked at networks built by the fashionable method of *preferential attachment* [318, 319, 320, 321]. The parameters of this process are N , the total number of nodes in the network, and m , the number of edges brought by each node as it is added to the network.

In ZMW, it is claimed that the fidelity curves are always concave; however, as we can see in Fig. 10.1, this is not so. Slonim and Tishby [322] state that for agglomerative information bottleneck (the algorithm employed by ZMW), concavity is an empirical result. (In a different, though closely related implementation of the information bottleneck concept, the fidelity curve is concave by construction.) From the numerical indications I've had so far, I think this is what's happening: the marginal fidelity curve is nicely monotonic if you plot it as a function of *the number of modules in the description*. If you change the horizontal axis to *the Shannon entropy of the description variable*, the marginal fidelity curve gets a bump.

All the stochastic processes we've considered in this thesis have been *classical*: nowhere have we incorporated quantum effects. Still, it is worth asking whether our multiscale structure formalism can be applied in quantum theory. How should one go about constructing a quantum analogue for our classical calculations? The first instinct might be to replace the Shannon index with the von Neumann entropy, the standard information measure of quantum physics. This choice is problematic, because while the von Neumann entropy is strongly subadditive, it is not monotonic [323], and

so it does not satisfy our basic axioms. Before we consider modifying those axioms, however, there is another possibility, one which might yield some conceptual insight.

In quantum mechanics, we are always calculating probabilities. We get results like, “There is a 50% chance this radioactive nucleus will decay in the next hour.” Or, “We can be 30% confident that the detector at position X will register a photon.” But the nature and origin of quantum probabilities remains obscure. Could it be that there are some kind of “gears in the nucleus,” and if we knew their alignment, we could predict what would happen with certainty? Fifty years of theorem-proving [324] have made this a hard position to maintain: quantum probabilities are more exotic than that.

But what we *can* do is reconstruct a *part* of quantum theory in terms of “internal gears” [325, 326, 327, 328]. We start with a mundane theory of particles in motion or switches having different positions, and we impose a restriction on what we can know about the mundane goings-on. The theory which results, the theory of the knowledge we can have about the thing we’re studying, exhibits many of the same phenomena as quantum physics. It is clearly not the whole deal: For example, quantum physics offers the hope of making faster and more powerful computers, and the “toy theory” we’ve cooked up does not. But the toy theory can include many of those features of quantum mechanics that are commonly deemed “mysterious.” In this way, we can draw a line between “surprising” and “truly enigmatic,” or to say it in a more dignified manner, between *weakly nonclassical* and *strongly nonclassical*. Results which are weakly nonclassical by this standard include quantum teleportation, quantum key distribution, the no-cloning theorem, coherent superpositions turning to incoherent mixtures by becoming entangled with the environment, quantum discord and many more.

The ancient Greek for “knowledge” is *epistēmē* (ἐπιστήμη) and so a restriction on our knowledge is an *epistemic restriction*, or *epistriction* for short [328]. Finding epistricted models for subtheories of quantum mechanics illuminates the question of what resources are required for quantum computation [329]. In addition, it suggests a way to apply our multiscale structure formalism directly, if not to the full variety of quantum phenomena, at least to an interesting and important subset. We can simply treat the states of the epistricted theory as probability distributions and use the Shannon index.

10.2 Category Theory for Moment Closures

A Petri net specifies a symmetric monoidal category [330]. Each truncation of the moment-dynamics hierarchy for a system yields a Petri net, and so successive truncations of the moment-dynamics hierarchy yield mappings between categories. Going from a pair approximation to a mean-field approximation, for example, transforms a Petri net whose circles are labelled with pair states to one labelled by site states. Category theory might be able to say something interesting here. Anything which can tame the horrible spew of equations which arises in these problems would be great to have. Ought we be considering, say, the strict 2-category whose objects are moment-closure approximations to an ecosystem, and whose morphisms are symmetric

monoidal functors between them?

10.3 Games with Variable Numbers of Players

When we discussed evolutionary game theory, we treated the size of social groups as constant. For example, in the Volunteer's Dilemma of §4.1, groups of $k \geq 2$ individuals coalesced out of the population, played the game, gathered their payoffs and fell back into the mix. In the network-structured version of the Volunteer's Dilemma, the size of each group of interacting players was fixed by the graph degree. But what if the number of players in a group changes from one interaction to the next? What if some contributions to an individual's fitness come from playing with one partner, some from playing with two and so on?

We can begin to answer this by importing technology from statistical physics. In chapter 5, we deduced the linked-cluster theorem, which states that for a certain way of assigning numerical values to graphs, the generating function over all graphs is the exponential of the generating function for connected graphs. We developed this in the context of probability distributions for random variables, where connected diagrams stand for cumulants, and the generating function over all moments is the exponential of that over cumulants:

$$\mathcal{Q}(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \langle x^n \rangle = \exp \left(\sum_{l=1}^{\infty} \frac{\langle x^l \rangle_c}{l!} z^l \right). \quad (10.4)$$

Now, it is time to deploy this machinery in evolutionary game theory. Let the weight of a k -vertex connected graph be the effect upon an individual's fitness due to interacting with k other agents at once. That is, the weight of a k -vertex connected graph is determined by the payoff of a $(k+1)$ -player game. If a focal agent plays a game with k_1 partners, then a game with k_2 partners and so on, we represent this by a graph comprising a k_1 -cluster, a k_2 -cluster and so on. If game payoffs are mapped exponentially to fitnesses, then the total fitness effect due to a sequence of accumulated payoffs is the product of the fitness effects due to each payoff in the sequence. That is, *the weight which we should assign to a graph comprising multiple disjoint clusters is the product of the weights of those clusters.*

What is the meaning of a sum over graphs in this context? One reason to add up a set of fitnesses is to find their average. If each possible sequence of cluster sizes satisfying $k_1 + k_2 + \dots + k_n = m$ is given equal weight, then the expected fitness effect due to a life history involving a total of m other organisms is proportional to the sum over all the appropriate graphs.

We have graph weights which compose in the proper manner. Therefore, we can use the linked-cluster theorem, Eq. (5.138), to deduce that the generating function over expected fitnesses for complicated life histories is the exponential of the generating function for the fitness effects of individual games.

10.4 Composition and a Multiscale Doi Formalism

In Chapter 7, we saw that classical stochastic dynamics can be treated with notations and tools adapted from quantum theory, specifically, raising and lowering operators (and associated entities like coherent states). The essential point was that these operators satisfied the commutator relation

$$[a, a^\dagger] = 1, \quad (10.5)$$

and, we noted, one implementation of this abstract algebra is differentiation and multiplication by a formal variable. This provides an instance of an unusual turnabout: mathematicians solve probability problems using generating functions, the *specific* representation, while the physicist approach uses the *abstract* algebra!

One thing which the specific representation suggests that the abstract algebra does not is the *composition* of functions: with two functions $f(z)$ and $g(z)$, it is easy to imagine evaluating g at a particular value and then plugging the result into f . We studied this construction in §5.6, where we saw that the linked-cluster theorem is a specific example of the iterated chain rule for differentiation. The composition of generating functions is used to great effect in combinatorics, as we shall now illustrate.

Let's say we have a set with k elements, and we wish to make some arrangement of those pieces. Perhaps we want to impose a linear order on the elements, or fashion them into a rooted tree, or make of them a point-set topology. In very general terms, we'd like to know how many ways we can do this: given some type of mathematical pattern F , how many ways can we implement F on a k -element set? Denote the set of all F -type arrangements on k elements by F_k ; then the number of ways to implement F on that many elements is the cardinality $|F_k|$. We can often deduce much about F from the generating function

$$|F|(z) = \sum_{n=0}^{\infty} \frac{|F_n|}{n!} z^n. \quad (10.6)$$

For example, the *vacuous structure* is just the arrangement of “being a finite set”; its generating function is e^z . Closely related is the *uniform structure*, which is like the vacuous structure except that it cannot be put on the empty set. Its generating function is therefore $e^z - 1$.

Consider the set F_k of F -type structures built out of a k -element set. If each such possible structure is weighted equally, then the “information content” in the statement “there's a structure of type F on this set of k elements” is the logarithm of $|F_k|$. This ties the combinatorial notion of “structure” to the informational one we used in Chapter 2.

The bars on $|F|(z)$ in Eq. (10.6) suggest that we will regard the whole generating function as the cardinality of some object. What kind of entity is the combinatorial species F ? For each k -element set, we have floating nearby the set F_k of F -type combinatorial arrangements we can perform on it; note that we care only about the cardinality of the set we are arranging, not about the character of its entries. Any two

sets of the same size—that is, any two finite sets between which exists a bijection—are equivalent for these purposes. We thus find ourselves manipulating $\mathbf{FinSet}_0$, the category whose elements are finite sets and whose morphisms are bijections. The combinatorial species F involves relating elements of $\mathbf{FinSet}_0$ with finite sets of possible arrangements, *i.e.*, with objects in the category $\mathbf{FinSet}$. Therefore, we understand F as a functor between these categories. Specifically, F send sets of structures to their substrate sets, being a “forgetful functor” which forgets the arrangement made out of the set’s elements. That is, a combinatorial species F can be represented

$$F : \mathbf{FinSet} \rightarrow \mathbf{FinSet}_0. \quad (10.7)$$

Investing in category theory pays off when we seek relationships among combinatorial species. Most importantly for our current purposes, we can compose two functors F and G to make a new species which encodes the “superstructure” of making an F -type arrangement from G -type structures. For example, we could construct a linear order out of trees. (Or we could bake a chocolate-chip cookie in which the chips are pieces of Oreo—or of Hydrox, which is equivalent up to isomorphism.) The generating function for the functorial composition of two species is the composition of the original species’ generating functions.

The Bell numbers $\{B_n\}$ are the number of ways to partition a set of n elements into nonempty subsets [228]. That is, B_n counts the number of ways to make a finite set out of nonempty finite sets, such that the total number of elements in the component sets adds up to n . We recognize this as making a uniform structure out of uniform structures, so we can say immediately that the exponential generating function for the Bell numbers is

$$B(z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n = e^{e^z - 1} - 1. \quad (10.8)$$

Note that in Eq. (10.8), we have a generating function which is (almost) equal to the exponential of a quantity which is itself (almost) an exponential of the formal variable z . This is more than slightly reminiscent of the linked-cluster theorem, Eq. (5.138). If b_l were to equal 1 for all l , the resemblance would be even stronger, which suggests we ought to look into possible generalizations of the composition which gave us the Bell numbers. We also have those -1 terms in Eq. (10.8), which came from choosing the uniform combinatorial species instead of the vacuum. The difference between the uniform and the vacuum species is in this case more than an off-by-one error: if we had tried to use the vacuum species instead, the number of possibilities would have blown up, as we could have interleaved an arbitrary number of empty sets into our arrangement. Composing with the vacuum species requires something more general and robust than ordinary combinatorial species [259, 331] which we shall explore below.

One problem with bringing partition functions like Eq. (5.138) into combinatorics and category theory is that we will sooner or later want graph weights which are not integers. Our graphs could be weighted by amounts of information content, interaction energies, probabilities or even probability amplitudes; even setting aside the quantum mechanics, we will have to weight graphs with real or at the very least rational numbers.

We require, then, a mathematical object whose cardinality is nonintegral.

One candidate solution is provided by the surreal numbers [213, 332, 333]. Joyal observed that the surreals, defined by Conway in terms of combinatorial games, can be formed into a category we could call **Game**, whose objects are game positions and whose morphisms are strategies [334]. By imposing a few extra conditions on the allowed game positions to eliminate infinite, infinitesimal and “pseudo-number” quantities, we can construct a category **RealGame** whose decategorification is precisely the real numbers $\mathbb{R}$.

Another approach, whose basic ingredients have been more fully developed in the literature, is to replace finite sets with *groupoids*, categories in which all morphisms are invertible. (In a groupoid with one object, the morphisms form a group.) Baez and Dolan [259] give the following formula for the cardinality of a groupoid: for each isomorphism class of objects within the groupoid, we pick a representative item and take the reciprocal of its number of automorphisms. The cardinality is then the sum over isomorphism classes.

$$|\mathcal{G}| = \sum_{[x]} \frac{1}{\text{aut}(x)}. \quad (10.9)$$

When all the morphisms in the groupoid are identity morphisms, groupoid cardinality reduces to set cardinality.

When we wrote the generating function for a combinatorial species, Eq. (10.6) we considered the preimage F_n of an n -element object in **FinSet**₀, which was a set of F -type structures living in **FinSet**. If the preimage of an object in **FinSet**₀ lived instead within a groupoid $\mathcal{G}$, its cardinality would not be constrained to the natural numbers, and could be a nonnegative rational quantity.

Take a groupoid $\mathcal{G}$ and make the groupoid **X** of $\mathcal{G}$ -colored finite sets. Then the functor

$$\Phi : \mathbf{X} \rightarrow \mathbf{FinSet}_0 \quad (10.10)$$

is a generalized species known as a stuff type. Morton [331] shows that the cardinality of $|\Phi|$ is

$$\begin{aligned} |\Phi| &= \sum_{n \in \mathbb{N}} |X_n| z^n \\ &= \sum_{n \in \mathbb{N}} |\mathcal{G} // S_n| z^n \\ &= \sum_{n \in \mathbb{N}} \frac{|\mathcal{G}|^n}{n!} z^n \\ &= \exp(|\mathcal{G}|z). \end{aligned} \quad (10.11)$$

Here, S_n is the permutation group on n elements, and $//$ denotes the “weak quotient” defined as follows:

Take a category C and a group G . The *strict action* of G on C is a map from group elements to functors. For every $g \in G$, the action A gives a functor $A(g) : C \rightarrow C$

satisfying $A(1) = \text{Id}_C$ and $A(gh) = A(g)A(h)$. The *weak quotient* of C by G is a groupoid, denoted $C//G$, whose objects are objects of the category C and whose morphisms are defined using the action A .

The categorical interpretation of Eq. (5.138) is that we are taking the cardinality of the functorial composition of two stuff types, one for “being a finite set” and the other encoding the graph-weighting procedure.

What does all this have to do with the Doi technology for stochastic processes? To make the connection, we recall that the Doi formalism is an abstract approach to probability generating functions. So, the question becomes, what is the meaning of the composition of two generating functions whose coefficients are probabilities instead of set cardinalities?

One application of generating functions in probability theory is to problems involving *randomly-sized sets of random variables*. Let N be a random variable taking nonnegative integer values, and define

$$Y = \sum_{i=1}^N X_i, \quad (10.12)$$

where each X_i is a random variable defined by a probability distribution p_X . That is, we pick some number n in accord with the distribution p_N , and then we draw n times from p_X and add the results up. For example, we could roll a 20-sided die, obtain the result 14, and then roll a 6-sided die 14 times. What is the probability distribution for the *total*?

The generating function for the random variable Y is the composition of those for N and for X . Specifically,

$$G_Y(z) = G_N(G_X(z)). \quad (10.13)$$

Suppose that, instead of being the outcome of a die roll, X is the number of objects present in a box. Then, naturally, N is the number of boxes, and Y is the total number of objects in all the boxes. Using the Doi method, we could write two vectors

$$|\phi_X\rangle = \sum_{i=0}^{\infty} p_X(i) |i\rangle, \quad (10.14)$$

$$|\phi_N\rangle = \sum_{i=0}^{\infty} p_N(i) |i\rangle. \quad (10.15)$$

Their composition is a third vector

$$|\phi_Y\rangle = \sum_{i=0}^{\infty} p_Y(i) |i\rangle \quad (10.16)$$

$$= \sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{d}{dz} \right)^i G_N(G_X(z)) \Big|_{z=0} |i\rangle. \quad (10.17)$$

If either p_N or p_X changes over time, then $|\phi_Y\rangle$ will change as well. This is a way of defining a *two-level* stochastic process. It is not, in its current form, the most useful approach, because the way we have set things up means that the dynamics of the two levels happen *independently*. But perhaps there is a way to make the levels interact: for example, for a box containing a large number of objects to split into a pair of boxes.

Composing a probability generating function with *itself* is useful in the study of iterated processes. If each of the objects inside a box is itself a smaller box, then Y is the total number of small boxes. Or, if the random variable X is the number of children produced by one individual, then Y is the number of grandchildren born to the N siblings. Setting $p_N = p_X$ amounts to declaring that the reproductive outcomes of one generation are the same, statistically, as those of the next.

This is a very different way of thinking about reproduction probabilities than we used in Chapter 7. Perhaps the intersection of the two can yield something interesting.

10.5 Multiplayer Games and Biodiversity Indices

In Chapter 5, we developed the Shannon index as a measure of information for a probability distribution. We started with the idea that, if p_i is our probability for seeing the i^{th} possible outcome of an experiment, then if we repeat the experiment N times, the least surprising number of times that outcome can occur is $N_i = Np_i$. A typical illustration of this is the experiment of plucking a letter at random from English-language text. Finding the letter E would be a less surprising event than coming up with a Q .

This is fine as far as it goes, but it does leave something out. The letters E and Q are, for many purposes, more dissimilar than, for example, S and Z , or F and V . In the first case, we have a consonant and a vowel, while the latter two pairs are all consonants. Furthermore, the latter two pairs are analogous: S and Z both stand for sibilants, one of which is voiced (that is, the vocal cords vibrate when the sound is articulated). Likewise, F stands for a voiceless consonant, and V for its voiced counterpart. From a phonetic standpoint, the event of encountering an F is more like the event of encountering a V than the event of finding an E is like that of meeting a Q . A quantitative measure of similarity would, justifiably, assign comparable values to the F – V and S – Z pairs, and a lower number to E – Q . Moreover, this is a separate question from how common or how typical the individual letters are.

For some purposes, the similarity between characters, and thus between the events of observing those characters, is a result of their evolutionary history. Our letters T and N have been distinct for as long as they have existed, and in a sense, longer than that: their ancestors in the Proto-Canaanite alphabet were distinct as well [335]. In contrast, the letter I did not diverge from J , or V from U , until after the Renaissance; and indeed dictionaries listed words beginning with U and with V together as recently as 1837 [336].

As with linguistic evolution, so with biological.¹ There is a meaningful sense in which

¹The analogy between these two processes was already clear to Darwin, who addressed it in Chapter 13 of the *Origin*. Later, he commented, “The formation of different languages and of distinct

the events of encountering two species from the same genus are more similar events than those of finding specimens from separate genera.

So, it is useful to consider information functions for cases in which we have not only a probability distribution over events, but a notion of similarity or contrast between those events as well [338].

Even if we do not have a full evolutionary history behind each type of entity we may encounter, we can still tabulate the characteristics of those entities. Suppose, for concreteness, that there exists a pool of n possible attributes, and an entity is defined by choosing exactly k of them. The attribute sets of any two entities may be disjoint, or they may have elements in common. If the attribute sets of two entities share common elements, then the events of encountering those entities are similar. But note that if we say “point” and “line” instead of “attribute” and “entity,” we have again the type of finite geometries we studied in Chapter 2. This suggests a new wrinkle: there are, very naturally, higher-order kinds of similarity, which cannot be deduced from lower-order relationships. Even if any two lines meet in a common point, a set of three lines might converge at a single intersection, or they might not.

Back in Chapter 5, we derived the Shannon index, which, we saw, reflects the extent to which a probability distribution is “spread out”: the Shannon index is maximized for a uniform distribution, and it attains its minimum value of zero when the distribution is a delta function. Another way to quantify the spread of a probability distribution is an *effective number*. This is a type of quantity, useful in mathematical ecology, which we can motivate with the following scenario.

Imagine that we have an urn full of marbles, and we presume that when we draw a marble from the urn, no choice is preferred over any other. If the urn contains N marbles, our probability of obtaining any individual one of them is $1/N$. But what if our probability distribution is not uniform, as it would be if we thought the drawing was rigged in some way? In that case, we can label the marbles with the integers from 1 to N , and we say that our probability for obtaining the one labeled i is $p(i)$.

We draw one marble, replace it and repeat the drawing. What is the probability that we will draw the same marble both times? Let the result of the first drawing be j . Then our probability for obtaining that marble again is $p(j)$, and to find the overall probability for drawing doubles, we average over all the choices of j :

$$p(\text{doubles}) = \sum_j p^2(j). \quad (10.18)$$

For a uniform distribution, this is

$$\sum_j p^2(j) = \sum_j \left(\frac{1}{N}\right)^2 = N \left(\frac{1}{N}\right)^2 = \frac{1}{N}. \quad (10.19)$$

That is, if all draws are equally probable, then the probability of a coincidence is the reciprocal of the population size. Turning this around, we can say that whatever our

species, and the proofs that both have been developed through a gradual process, are curiously the same” [286]. It is an area of ongoing conceptual exchange [337].

probabilities for the different draws, the effective size of the population is

$$N_{\text{eff}} = \left[\sum_j p^2(j) \right]^{-1}. \quad (10.20)$$

So far, we have presumed that any pair of outcomes is as good as any other. For some problems, this can be good enough. However, if we are trying to find the effective number of organisms present in an ecosystem, we must face the fact that some pairs of species are more closely related than others.

Leinster and Cobbold have proposed a framework for biodiversity indices which systematizes and extends many prior developments in that field [339]. The basic input data they consider is a set of probabilities $p(i)$, which characterize the relative preponderances of species in an ecosystem, and a *similarity metric* which indicates how closely species i resembles species j . Their diversity indices depend on a *sensitivity parameter*, call it q , which indicates the relative emphasis placed on rare species versus common ones. The larger one makes q , the less sensitive the diversity index is to improbable species.

The Leinster–Cobbold diversity index is

$${}^q D^{\mathbf{Z}}[\mathbf{p}] = \left[\sum_i p(i) \left(\sum_j Z_{ij} p(j) \right)^{q-1} \right]^{\frac{1}{1-q}}. \quad (10.21)$$

Here, we follow Leinster and Cobbold in writing $\mathbf{Z}$ for the matrix of similarity values Z_{ij} . If $Z_{ij} = \delta_{ij}$, we recover the case in which distinct species are considered wholly unrelated to one another, which is often (and unrealistically [338, 340]) assumed in much of the older work on biodiversity. We can motivate Eq. (10.21) in the following way: if Z_{ij} is the similarity between species i and species j , then summing $Z_{ij} p(j)$ over all possible values of j will give the “ordinariness” of species i . The “average ordinariness” of the whole ecosystem is then just the mean of this taken over the probability distribution $p(i)$. Because “diversity” ought to be inversely related to average ordinariness, we take the reciprocal. Eq. (10.21) generalizes this to *power means* of order $q - 1$. Most of our attention will be focused on the special case $q = 2$; that is, we will mostly consider the index defined with the ordinary arithmetic average.

Biodiversity indices allow us to compare ecosystems. For example, given two test tubes full of microorganisms, one which we characterize by a probability distribution $\mathbf{p}$ and the other which we characterize by $\mathbf{r}$, we can compare the “diversity profiles” of the two microbial ecosystems by computing ${}^q D[\mathbf{p}]$ and ${}^q D[\mathbf{r}]$ as functions of q [339]. We can delve more deeply if the two ecosystems are composed of the same species, that is, if $\mathbf{p}$ and $\mathbf{r}$ are not just of the same length, but defined over the same events. Suppose that we pipette a microbe at random from the first test tube, and that we identify this microbe as being of type i on our list of all possible microbe varieties. What is the probability that the act of pipetting out a random microbe from the *second* test

tube will yield an organism of the *same* type? By definition, it is $r(i)$. What, then, is the *average* probability over *all possible microbe varieties* that the results of the two experiments will “collide” in this way? The answer is the expectation value of $r(i)$ with respect to the probability distribution $\mathbf{p}$, or in other words, the dot product of the two probability vectors $\mathbf{p}$ and $\mathbf{r}$. As before, we would like a measure of diversity to be inversely related to a tendency toward coincidence. We can therefore define a *cross diversity* as

$${}^2D[\mathbf{p}; \mathbf{r}] = \frac{1}{\sum_i p(i)r(i)}. \quad (10.22)$$

We can also include a distance metric $\mathbf{Z}$ in the cross diversity, as we had with the Leinster–Cobbold index:

$${}^2D^{\mathbf{Z}}[\mathbf{p}; \mathbf{r}] = \frac{1}{\sum_{ij} Z_{ij}p(i)r(j)}. \quad (10.23)$$

As before, this reduces to the ordinary cross diversity of Eq. (10.22) in the limiting case that Z_{ij} is a Kronecker delta, that is, when a species resembles only itself and is distinct from all others.

For reasons stemming from biology, Leinster and Cobbold prefer to set up their indices as “effective numbers” of species present, rather than as entropies. This has become a standard practice in mathematical ecology. In just about any circumstance, it’s reasonable to say that an island with four equally abundant, unrelated species is only half as biodiverse as an island with eight equally abundant, unrelated species. Effective numbers preserve this desirable scaling property, while entropies of the forms familiar from information theory do not [341]. Relating these effective numbers to various entropies people have defined is not conceptually difficult. Generally, we expect an entropy to be logarithmically related to an effective-number diversity measure, because an entropy should count the number of questions needed to specify an item in a set. (For example, in the study of complex networks or graphs, the “effective degree” of a vertex is the exponential of the entropy of its edge-weight distribution [342].)

We can think of these diversity indices in another way, which suggests a natural generalization. The key comes from *game theory*. Imagine a game in which each player’s goal is to match the move made by the other. The score earned by a player who makes move i is 1 if the other player also makes the move i , and 0 otherwise. If the players make their moves *randomly*, in a way characterized by the probabilities $p(i)$, then the *expected payoff* obtained by either player is just $\mathbf{p} \cdot \mathbf{p}$. Now, suppose the matching is not such an all-or-nothing affair: perhaps there are wild cards in the deck, so that the ace ($i = 1$) can match any other. Or, perhaps matching one card with another of the same suit is almost as good as matching it with a copy of the same card. Then the expected payoff will include cross terms, since the score of an i matched against a j is no longer just δ_{ij} . Diversity indices, then, are *measures of expected welfare in games whose goal is agreement*.

This new perspective hints at a generalization: what about games played by more than two players?

Leinster and Cobbold think in terms of *distance* between species, and distance is

naturally a pairwise thing. However, as they point out, some of the earlier work in the area considered *inter-species conflict* instead [343]. Conflicts, or interactions more generally, do not have to break down into pairwise relationships. In human affairs, what Alice does in the company of Bob and Carol does not have to be a linear combination of what she does when alone with Bob and when alone with Carol. In game theory, the payoff in a multiplayer game is not restricted to being a linear sum of pairwise games (recall Chapter 4). Or, consider a parasite species with multiple hosts in its life cycle: the total effect on humanity due to *Anopheles* mosquitos and *Plasmodium* microbes depends on both species taken together.

We imagine, therefore, an “interaction tensor” of the form Z_{ijk} , which tells us how the presence of species i and j taken in combination affects a focal species, k . The natural modification of Eq. (10.21) is

$${}^q D^{\mathbf{Z}}[\mathbf{p}] = \left[\sum_k p(k) \left(\sum_{ij} Z_{ijk} p(i)p(j) \right)^{q-1} \right]^{\frac{1}{1-q}}. \quad (10.24)$$

Biodiversity measures, generalized to multiplayer games as we have done here, have an application in quantum information theory, of all places. This is one of the oddest bits of interdisciplinary boundary-jumping which I have seen, so I think it’s worth talking about for a while.

In quantum physics, we take what we think we know about a system, roll it into a density operator ρ , and use that density operator to make statistical predictions about what the system might do in particular experiments. But presenting that information as a matrix operator is not always the most illuminating choice. We can actually rewrite any finite-dimensional density matrix as a probability distribution, using the idea of *informationally complete measurements* [344, 345]. These are generalized measurement procedures (positive operator valued measures, or POVMs) which have an appealing ability: given a probability distribution over the possible outcomes of an informationally complete POVM, we can compute all the statistics which we could have gotten using the density matrix. Such POVMs can be constructed in any finite-dimensional Hilbert space [224, 346]. The nicest variety are the *symmetric* informationally complete POVMs, known familiarly as SICs [347, 348]. A SIC for a d -dimensional Hilbert space is a set of d^2 operators $\{E_i = \frac{1}{d}\Pi_i\}$ where the rank-one projection operators $\{\Pi_i\}$ satisfy

$$\text{Tr}(\Pi_k \Pi_l) = \frac{d\delta_{kl} + 1}{d + 1}. \quad (10.25)$$

An arbitrary density matrix ρ can be decomposed in terms of SICs. If $p(i)$ is the probability that performing a SIC measurement on the system yields the outcome

labeled by i , then

$$\rho = \sum_{i=1}^{d^2} \left((d+1)p(i) - \frac{1}{d} \right) \Pi_i = (d+1) \sum_{i=1}^{d^2} p(i) \Pi_i - \mathbb{I}. \quad (10.26)$$

Exact expressions for SICs have been found for dimensions 2–16, 19, 24, 28, 31, 35, 37, 43 and 48 [349]. High-precision numerical approximations have been discovered for dimensions 2–67 [350], and more recently, E. Schnetter has claimed numerical solutions for dimensions 68–76, 78–81, 83–85, 87, 89, 93 and 100 [349]. It is not known whether SICs exist for all values of d , but it has become commonplace to assume that they do.²

The extremal states in the space of density matrices are the “pure” states, which satisfy the condition $\rho^2 = \rho$. Thanks to a theorem of Flammia, Jones and Linden [352, 353, 354], we can also characterize pure states as those Hermitian matrices satisfying

$$\text{Tr} \rho^2 = \text{Tr} \rho^3 = 1. \quad (10.27)$$

This result is well worth calling a remarkable theorem: it is simple, powerful and easy to prove once asserted, but it was apparently missed completely until 2004 [211]. In turn, this definition of purity yields the following two conditions on the probability distribution $p(i)$ [349, 355, 356]. First,

$$\sum_{i=1}^{d^2} p(i)^2 = \frac{2}{d(d+1)}, \quad (10.28)$$

and second,

$$\sum_{ijk} c_{ijk} p(i)p(j)p(k) = \frac{d+7}{(d+1)^3}, \quad (10.29)$$

where we have defined the real-valued, symmetric three-index tensor

$$c_{ijk} = \text{Re Tr}(\Pi_i \Pi_j \Pi_k). \quad (10.30)$$

The second condition, Eq. (10.29), has been nicknamed the *QBic equation* [211]. The full state space is the convex hull of probability distributions which meet Eqs. (10.28) and (10.29). It would be interesting to be able to motivate these equations from something other than the pre-existing quantum formalism: is there a reason, independent of quantum physics, to care about functionals of probability distributions like we see on the left-hand sides of Eqs. (10.28) and (10.29)?

Formally, the conditions defining quantum-state purity, Eqs. (10.28) and (10.29), become

$${}^2D^\delta[\mathbf{p}] = \frac{d(d+1)}{2} \quad (10.31)$$

²Following a conjecture about a way to reduce the number of equations which must be solved to obtain a SIC [211, 351], I’ve found low-precision evidence that such a structure exists in $d = 77$ as well.

and

$${}^2D^c[\mathbf{p}] = \frac{(d+1)^3}{d+7}. \quad (10.32)$$

This form of the QBic equation is, of course, still a cubic constraint, in that it employs three instances of the probability vector $\mathbf{p}$. However, as far as the sensitivity parameter q is concerned, it is a “second order” equation, as we have kept to the special case $q = 2$.

What can we learn from these restatements of the pure-state conditions? Amusingly, assigning a pure state to a quantum system means that the *effective number* of possible outcomes for a SIC experiment which one is willing to contemplate is just $\binom{d+1}{2}$. The fact that an effective number works out to be a combinatorial quantity hints strongly, at least to me, that this is a promising avenue to explore: combinatorics is, after all, the art of counting cleverly. (Thinking in terms of effective numbers has at least a few pence of “cash value” [211], in that it means I can remember what goes on the right-hand side of the quadratic purity condition.) Another way to think of this is that when all SIC outcomes are judged as equiprobable, that is to say $p(i) = \frac{1}{d^2}$, the effective number of experimental outcomes is the total number which comprise the SIC: ${}^2D^\delta = d^2$. So, if we focus on the quadratic constraint, ascribing a pure state means neglecting $\binom{d}{2}$ possible outcomes of a SIC experiment. Entertainingly, this is also the best known upper bound on the number of entries which can be zero in a quantum-state assignment $\mathbf{p}$ [356]. This is not a coincidence: we can deduce that bound by starting with the normalization of $\mathbf{p}$ and squaring to find

$$\left(\sum_i p(i) \right)^2 = 1. \quad (10.33)$$

We then apply the Cauchy–Schwartz inequality to find, writing n_0 for the number of zero-valued entries in $\mathbf{p}$,

$$(d^2 - n_0) \sum_{\text{nonzero}} p(i)^2 \geq \left(\sum_{\text{nonzero}} p(i) \right)^2 = 1. \quad (10.34)$$

We see the inverse of the quadratic diversity appearing on the left-hand side. Consequently,

$$n_0 \leq d^2 - {}^2D[\mathbf{p}], \quad (10.35)$$

and from Eq. (10.28) we know the right-hand side equals $d(d-1)/2$, as advertised.

This bound is perhaps not the tightest possible. In fact, it is suspected [349] that the actual upper bound on the number of zeros permitted in a quantum state is just the dimension, d .

In addition, quantum mechanics implies a constraint on *pairs* of probability vectors [349, 356, 357]. If we begin with state assignments written as density matrices ρ and σ , then using Eq. (10.26) we can deduce that the Hilbert–Schmidt inner product of

those state assignments is

$$\text{Tr } \rho\sigma = d(d+1) \sum_i r(i)s(i) - 1. \quad (10.36)$$

Because the Hilbert–Schmidt inner product is always nonnegative, two quantum state assignments defined on Hilbert spaces of the same dimension d “can never be too nonoverlapping” [349]:

$$\sum_i r(i)s(i) \geq \frac{1}{d(d+1)}. \quad (10.37)$$

Comparing this to Eq. (10.22) above, we have a bound on the cross diversity of $\mathbf{r}$ and $\mathbf{s}$:

$$^2D[\mathbf{r}; \mathbf{s}] \leq 2 \binom{d+1}{2}. \quad (10.38)$$

From the first purity condition, Eq. (10.28), we know that all valid probability vectors lie within a ball. The dot product of any two state assignments will therefore obey the bound

$$\sum_i r(i)s(i) \leq \frac{2}{d(d+1)}, \quad (10.39)$$

meaning that their cross diversity is also bounded from below:

$$\binom{d+1}{2} \leq ^2D[\mathbf{r}; \mathbf{s}] \leq 2 \binom{d+1}{2}. \quad (10.40)$$

The lower bound is saturated if $\mathbf{r} = \mathbf{s}$ and $\mathbf{r}$ is a pure state.

Tabia [358, 359] has discovered a fascinating simplification of the QBic equation, Eq. (10.29), in the case of a qutrit, a system whose Hilbert space has dimension $d = 3$. In this special case, Eq. (10.29) can be reduced through a clever choice of SIC to

$$\sum_i p(i)^3 - 3 \sum_{(ijk) \in S(9)} p(i)p(j)p(k) = 0. \quad (10.41)$$

Here, $S(9)$ denotes the *Steiner triple system* of order 9, a set of 12 elements which can be found by cyclically tracing all the horizontal, vertical and diagonal lines in the array

$$\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array} ; \quad \text{that is to say,} \quad S(9) = \begin{array}{ccc} (123) & (456) & (789) \\ (147) & (258) & (369) \\ (159) & (267) & (348) \\ (168) & (249) & (357) \end{array}. \quad (10.42)$$

The order-9 Steiner triple system also rejoices in the name of the *Hesse configuration* [358, 359, 360, 361]. This construction is also an example of a dual affine plane, a type of incidence geometry which is known to be relevant to SICs more generally [25, 27, 362], and which we saw all the way back in Chapter 2.

We consider “striations” of the 3×3 array: we can carve it up into horizontals, verticals, left-leaning diagonals or right-leaning diagonals, and each of these four striations divides the array into three parallel sets of numbers. That is, each striation produces one row of the table in Eq. (10.42).

We can easily check the simplified QBic equation, Eq. (10.41), in one case where we know it should hold true: for the SIC states themselves. In a SIC representation, the k^{th} vector making up the SIC has a $1/d$ in the k^{th} slot and $\frac{1}{d(d+1)}$ everywhere else. Specializing to $d = 3$, any of the nine SIC states appears in exactly four of the index triples listed in Eq. (10.42). This makes checking that all nine SIC states satisfy Eq. (10.41) a straightforward arithmetic problem.

Note that we can recast the $d = 3$ QBic equation, Eq. (10.41), as

$$\sum_i p(i)^3 + 3 \sum_{(ijk) \in S(9)} p(i)p(j)p(k) = 2 \left(3 \sum_{(ijk) \in S(9)} p(i)p(j)p(k) \right) = 2 \left(\sum_i p(i)^3 \right). \quad (10.43)$$

On the right-hand side of Eq. (10.43), we have the inverse of the ternary diversity we defined back in Eq. (10.24), with experimental outcomes treated as wholly dissimilar. On the left-hand side, we have something which is starting to look like a ternary diversity with some sets of outcomes distinguished as more similar than others.

Let Y_{ijk} be the completely symmetric tensor which is 1 if $(ijk) \in S(9)$ and 0 if no permutation of (ijk) is in $S(9)$. Then

$$\sum_{ijk} Y_{ijk} p(i)p(j)p(k) = 6 \sum_{(ijk) \in S(9)} p(i)p(j)p(k). \quad (10.44)$$

Now, let Z_{ijk} be 1 if all subscripts are equal and $\frac{1}{2}Y_{ijk}$ otherwise. Then the $d = 3$ QBic equation reads

$${}^2D^{\mathbf{Z}}[\mathbf{p}] = \left(\sum_i p(i)^3 + 3 \sum_{(ijk) \in S(9)} p(i)p(j)p(k) \right)^{-1} = \frac{1}{2} {}^2D^{\delta}[\mathbf{p}]. \quad (10.45)$$

For qutrits, pure states are those for which treating as similar the proper sets of SIC outcomes reduces the diversity of possible outcomes by 2.

The Hesse configuration has other interesting properties in relation to qutrit SICs, including a result I found about “mutually unbiased bases” and compatibility criteria for quantum-state ascriptions [363]. This is a result of quantum information theory—actually, a correction to others’ earlier work in that field—which I proved thanks to an interest in mathematical biology and complex systems. These calculations are, if taken just as mathematics, neutral on philosophical questions about quantum physics. From that perspective, they are matters of complex projective geometry. However, the research tradition they derive from, and which they might feed back into, is a philosophy which treats quantum physics itself in evolutionary terms, as a tool which agents immersed in the creative profusion of the world can use to make the best of

life's Darwinian contest [211, 349].

10.6 Gauge Theory and Evolution

The study of evolution has a certain conceptual intersection with the subject of economics. Both are concerned with the effects of limited resources. And, both make use of game theory. On the evolutionary side, however, we are not concerned with whether agents are “rational”; nor do we start with an assumption that a system is “efficient” or that the participants in it are by any standard well-informed. This is a significant difference in outlook between evolutionary theory and economics, and arguably, economics needs to catch up [31, 364, 365, 366, 367, 368].

I can think of one idea, though, which is rather in the borderlands of economics, which might fruitfully be transposed over into mathematical biology. That is the application of *gauge theory* and *differential geometry* to economic indices. This subject was inaugurated by Malaney [369], in collaboration with Weinstein. More recently, it has been promoted by Smolin [370] and discussed by Baez [371].

A big concern in the Malaney thesis is how to define a cost-of-living index when the goods which are relevant to daily life change over time. If Richard III regarded a horse as a fair trade for his kingdom, how much of England should we be able to swap for a Prius? We can answer this kind of question quantitatively, as long as old goods remain in circulation at least temporarily as new ones are introduced. During the period of overlap, we can evaluate the price at which agents participating in the economy will trade an old good for a new one and vice versa. By chaining overlap intervals, we can gradually eliminate all the goods in the initial “basket,” while maintaining an unbroken sense of what counts as a decent standard of living.

An evolutionary analogue would be the origin of new traits. A common mechanism for this is gene duplication: a whole stretch of DNA accidentally gets copied twice, so the offspring carries two copies of the same gene. This doesn't make much immediate difference, but it does provide redundancy: over the generations, mutations accumulate, and because mutations which knock out one gene leave the other intact, the species is more resilient. Over time, one of the gene copies can gain a new function, while the other keeps doing the old. This is studied in the context of gene interaction networks, where it's known as the “duplication-divergence model” [372].

Adapting the Malaney–Weinstein economics stuff might provide a way to talk about what “fitness remaining constant” means in this context, and to write equations for population dynamics.

To lay the groundwork for this project, we can establish a dictionary from economics to evolution. Instead of *goods and services*, we can speak of *traits*. In economics, one considers the *amount of a good* held by an agent; the counterpart in evolutionary dynamics is the *expression level* or *quantitative value* of a trait. Rather than a *pricing system* that assigns prices to baskets, we have *fitness functions*, given in the simplest case as a linear combination of trait values with coefficients (*e.g.*, the β values of Chapter 9).

The Malaney thesis [369] defines a “barter” as a set of debts and possessions whose

net monetary value is zero. For example, if six cupcakes are worth ten cookies, then a debt of ten cookies can also be paid by six cupcakes. The evolutionary analogue is a mutation which does not affect the reproductive fitness of the individual. On the molecular level, this happens all the time: replication errors swap out nucleotides so that the offspring end up carrying different DNA sequences, but those sequences still code for the same protein. In the mathematical setup we discussed in the previous chapter, a neutral mutation could be one trait diminishing in value while another one increases, so the w_i we get by taking the dot product doesn't change.

On the economic side, we have phenomena like inflation, in which the function from baskets to prices changes over time. This has a natural counterpart in evolution; for example, shifts in external environmental conditions can be represented as variations in the parameters of the fitness function.

By carrying over the basic notions of Malaney [369] to the new terminology, we have our first result: *If the covariant derivative of an evolutionary history vanishes, then that history consists of neutral mutations.*

So far, in fashioning this map from one jargon to another, we have neglected population structure. It would be interesting to push further, into models where the fitness function is not constant across the population. Moreover, the work on the economics side presumes that price scales linearly with quantity, and we know that doing mathematical biology means facing up to nonlinearity sooner rather than later.

Gauge theory provides a platform for understanding the phenomenon of moving through a circuit and not ending up exactly how you began. In Chapter 3, we saw an evolutionary dynamic which is reminiscent of this: a consumer strain can go from rare and successful to rare and dwindling, while local information indicates no difference between the environmental conditions. Can we express this phenomenon in terms of fiber bundles?

11 Conclusions

11.1 Review

Whatever universe a professor believes in must at any rate be a universe that lends itself to lengthy discourse.

—William James [2]

We began, many chapters ago, with the idea of a *complex system* being one which exhibits organization at multiple scales. Formalizing this concept mathematically, we saw that indices of multiscale structure can describe the patterns which result from evolutionary processes. Then, we saw how the emergence of multiscale structure can control the evolution of trait values in an ecosystem, and we explored how we can expand game theory to multiplayer scenarios, providing another angle on the question of scale. With tools from probability theory in hand, we brought adaptive dynamics into the stochastic regime, and we found a way to quantify the interplay between mutation and selection, revealing how the outcome depends upon population structure.

Academic writing traditionally patterns itself rather like the Twelve Labors of Herakles [373]. First, there is the statement of the problem, and then, a litany of failed attempts to solve it, which we call a “literature review.” Next, a new and better solution is proposed, and its triumph is proclaimed. I elected to deviate from this template, as far as the overall layout of this thesis is concerned. This choice hinged upon a tradeoff that I think is important enough to warrant discussing explicitly. A literature review is a specialized history in miniature, and so it is at this juncture worth thinking about historical expositions of science more generally.

This thesis has addressed several variations on the theme of multiscale structure in evolutionary dynamics. I have tried to make the material flow as smoothly as possible, and to build up the ideas in a sequence which helps them come across transparently. It may be, however, that in doing so, I have clouded the distinction between new research and old. The first glimpse of a discovery is not always the most clear, after all, and later treatments of a topic can have the advantage of experience over the early ones. Initial reports are commonly soaked in the confusions of the time, which require hard work (and, sometimes, as the proverb says, funerals) to overcome. Consequently, an exposition which aims to express the current synthesis is apt to be ahistorical.

Any field of science which has reached a sophisticated level of development is susceptible to this problem. Suppose I want to teach a classful of college sophomores the fundamentals of quantum mechanics. There is a standard “physicist’s history of physics” [374] which goes along with this, one that progresses through a familiar litany of famous names: Planck, Einstein, Bohr, de Broglie, Heisenberg, Schrödinger, Born.

11 Conclusions

We like to go back to the early days and follow the development forward, because the science was simpler in its initial stage—or so we tend to believe.

The problem is that all of these men were highly trained, professional physicists who were thoroughly conversant with the knowledge of their time. But this means that any one of them knew more classical physics than a modern college sophomore. They would have known Hamiltonian and Lagrangian mechanics, for example, in addition to techniques of statistical physics. Unless you know what they knew, you can't really follow their thought processes, and we don't teach big chunks of what they knew until after we've tried to teach what they figured out! For example, if you aren't fairly conversant with thermodynamics and statistical mechanics, you won't be able to follow why Planck proposed the blackbody radiation law he did [375], and a crucial step of the development will be lost, without your even knowing it.

Consequently, any “historical” treatment at the introductory level will probably end up conventionalized. One has to step extremely carefully! Strip the history down to the point that students just starting to learn the science can follow it, and you might not be portraying the way the people actually did their work. That's not so bad, as far as learning the facts and formulæ is concerned, but you open yourself up to all sorts of troubles when you get to talking about the process of science. Are we doing physics differently than folks did N or $2N$ years ago? If we aren't, is that a problem? Well, we sure aren't doing it like they did in the textbooks we learned out of.

Carelessly repeating a “scientist's history” instead of teaching a history of science leads to a kind of inadvertent myth-making that Stephen Jay Gould designated “text-book cardboard” [376]. An example from biology would be the assertion that no one put genetics and natural selection together until the 1930s [287]. In physics, to name one of many possibilities, we have the canard that in the last years of the 19th century, physicists thought that the only remaining task was to calculate answers to more decimal places.¹ The “physicist's history” of twentieth-century physics is replete with textbook cardboard [383, 384, 385, 386], and no doubt we'll keep this tradition going in the twenty-first.

Having argued that there exists a cause for concern, I am now in the unheroic position of admitting I have no way to solve it. The only way out seems to be flexibility, sacrificing one objective for another as the circumstances permit.

The field of complex-systems research has an additional challenge. Many of the models it employs are defined computationally: we begin with a specification of the model, and then we implement it as a computer program. Not infrequently, that implementation is within the range of a fairly novice programmer. The primary challenge

¹This sentiment is often attributed to Kelvin, but never with an actual pointer to a primary source.

In a 1900 lecture, expanded the next year to an essay, Kelvin described two “clouds over the dynamical theory of heat and light” [377]. Dispelling the first cloud turned out later to require special relativity, and removing the second was a task for quantum mechanics. Kelvin concludes his discussion of the first problem by the remark, “I am afraid we must still regard Cloud No. I. as very dense.” Figures as prominent as Tait [378], Gibbs [379] and Maxwell himself [380] all pointed out that classical physics fails to grasp the specific heats of gases. A text as widely admired and merchandised as the *Feynman Lectures* laid out the historical situation [381, 382]. Even so, the assertion of *fin-de-siècle* physicists' naïve folly lives on, helping thinkpieces to be glib and cocktail talk to be smug.

lies not in the coding, but in deciding which model would be interesting to explore, and in knowing how to investigate it systematically. The *mathematical* prerequisites for working with adaptive-network models, say, are less demanding than they are for many topics at the modern physics frontier. Furthermore, when analytical treatments are possible for complex-systems models, they often only apply to special cases. Nonequilibrium phase transitions provide a good example: a basic implementation of directed percolation is easy to code [74], but the statistical field theory which describes it is difficult to obtain (Chapter 7), and the elaborate mathematics is most useful near the critical point.

It may be that the special case of a problem, the case amenable to analytical work, was historically discovered first, and computational methods which made the general case accessible only came later. However, as we have discussed, computational methods can admit easier expositions than densely mathematical ones. Thus, an explanation which moves from the familiar to the esoteric would invert the historical order.

As a result, presenting all of the literature review first [387] would have been a dissatisfying scheme for this thesis. Moreover, such a sequence would have required muddling through the remarkably confused MLS/IF literature we surveyed in Chapter 9, before seeing the simulation results in Chapters 3 and 4, which are significantly easier to appreciate. Likewise, for a reader trained in physics, the Price equation we met in Chapter 9 is likely to be unfamiliar and, indeed, to appear arcane and even overwrought. In contrast, the dynamical stability analysis presented in Chapter 4 is a technique more routine to a physics student, albeit applied to a less common-or-garden problem. Again, choosing to begin with the approachable material means that we turn history on its head.

That said, optimizing towards one goal can impede the achievement of another. With this in mind, I will use this section to revisit the previous chapters and more cleanly separate research from review.

Chapter 2 discussed a general formalism for multiscale structure, based on information theory. We saw two indices of structure, the complexity profile and the Marginal Utility of Information (MUI). The complexity profile was introduced some years ago by Bar-Yam [4]. More recently, Allen developed an axiomatic foundation upon which the complexity profile and related quantities could be constructed. The three of us coauthored a paper on the subject [3]. During the course of that project, I derived the binomial transform equation for computing the complexity profile in the special case of exchange-symmetric systems, Eq. (2.37). I also worked out the illustrative examples using combinatorial geometry, which our three-author paper was already long enough without. The complexity profiles for the imitation dynamics (noisy voter model) and the frequency-dependent Moran process are new calculations for this thesis.

Chapter 3 is based on a paper by Gros, Bar-Yam and myself, as I indicated in its concluding note. The organism-swapping test and the use of survival-probability curve intersections were new contributions of that paper. In addition, the use of percolation arguments to find the critical thresholds in certain limits had not been done for that model before, and the discrepancy between predator-prey and epidemic dynamics (§3.3.6) had not been commented upon. Relating the correlation lengths to the evolved transmissibilities is new, as are the crossover in the perimeter-area

11 Conclusions

relationship and the 99th-percentile curves. The scaling argument which shows the inability of pair approximation to handle the correlation length is a new application of earlier semi-numerical work. So is the combination of the pair approximation with the coagulation and fragmentation model to predict how τ will evolve.

The next chapter, on multiplayer games, contains several novelties. I encountered the Volunteer's Dilemma in papers which argued that it was an understudied type of scenario [56, 201]. While writing Chapter 4, however, I found that I had to treat the game in ways which I hadn't read about. The continuous-time, well-mixed dynamical systems, with their baseline growth and death rates, logistic forms, pessimistic view of the cost of Volunteerism and unfixed total population size, are unusual. The lattice models are, likewise, atypical with respect to the literature. The analytical computations which conclude the chapter are an extension to multiplayer games of ideas which had only been applied to dyadic interactions.

Chapter 5 is largely a review of material which would have otherwise required references to selected chapters in a scattering of textbooks. The part about an evolutionary analogue of the Jeffrey rule is, to my knowledge, not covered elsewhere. (I've mentioned the possibility over the past few years to people better-read than I, and prior art never came up.) Anomalous cross-diffusion terms, as seen in §5.8, are known in the literature [132, 133], but that literature arrives at them in a much more complicated way.

I learned at school the physicist's justification of the Fokker–Planck equation [388] which now appears in Chapter 6, and I read about the deterministic limit of adaptive dynamics [230]. The applications to the continuous Prisoner's Dilemma and Snowdrift games are new developments for this thesis, as is the connection to random matrix theory.

Both Chapters 8 and 9 are pedagogical reviews, for the most part. The argument that the “modified mean-field model” is insufficient for eco-evolutionary considerations (§8.6) is my contribution. My approach to the interconversion between MLS-A and neighbor-modulated fitness calculations (§9.4) is more abstracted and generalized than Bijma and Wade's [46], which is where I began.

The work on which this thesis builds is, in many places, fairly new itself. This adds an extra layer to the challenge of exposition: the ideas are sometimes recent enough that their significance has yet to be fully hashed out. I hope this thesis can be a part of that process.

In 1996, John Horgan published a book titled *The End of Science*. This attracted some serious criticism at the time [389], but the passing of years has brought an even more strongly negative review. To wit: in 1997, the first *E. coli* genome was published [390], and in 1998, the expansion of the Universe was found to be accelerating [391, 392]. Accomplishments like these are not the final fits of a dying enterprise; nor are they the concluding, semi-senescent huzzahs before a quiet retirement. They are new beginnings, which raise up new opportunities and which all the work that comes after must acknowledge. These are the kind of discovery which go beyond providing answers: they change the questions which we are able to ask. And to these examples we could add many more.

I mention this because the bulk of the references on which I leaned most directly

date to the post-*End-of-Science* period. The complexity profile, for example, emerged in 2004 [4], and the axiomatic formalism of multiscale structure only last year [3]. Chapter 4's analytical calculations depend on a scheme which was likewise published in 2014 [37], and the Fokker–Planck treatment of stochastic adaptive dynamics is an extension of research reported the year before that [230]. The theory of the structure coefficient σ , which provides a convenient way to discuss many mathematical models of evolution in a unified fashion, was put forth in 2009 [231], and it is perhaps still not as well known as it should be.

11.2 Just One More Thing

Another way in which my organizational scheme is a little heterodox is that I decided to move the Acknowledgments to the end. Partly, this is to soothe my conscience: having all those pages come before now makes me a little more inclined to feel that this work is substantial, and thanking those who helped it along is not entirely damnation by faint praise. And, partly, this arrangement is to benefit those who, like me, skip to the end to see what the thing is about.

I am indebted to my collaborators, first of all to those with whom the relation is made official by coauthorship: Marcus de Aguiar, Benjamin Allen, Yaneer Bar-Yam and Andreas Gros. Many portions of this thesis are my attempts to build on something Ben has done. It would be a much slimmer document without his achievements as wellsprings. Without our conversations, it might not exist at all.

Special mention must also be given to my thesis committee: Yaneer Bar-Yam, Aparna Baskaran and Albion Lawrence. Their input throughout the stages of this project has been consistently helpful.

John Baez had very kind things to say about the work which became Chapter 3, and also stimulated parts of Chapters 5 and 7. Karla Z. Bertrand read many iterations of what is now Chapter 3, and in addition helped test my presentation in Chapter 2. Chris Fuchs got me to thinking about some of the matters which inform Chapters 5 and 10.

Credit must go to my far-flung correspondents, who kept my curiosity active and my sense of honesty engaged by reminding me of questions in varied fields of science. A large fraction of our interactions these past few years have taken place through a strange one-to-many telegraph.

@archymck
 @arikia
 @artologica
 @arutherfordium
 @AstroKatie
 @bengoldacre
 @biochembelle
 @bug_gwen
 @carlzimmer

11 Conclusions

@cgranade
@cryptogoth
@csferrie
@cuttlefishpoet
@dabacon
@Dhunterauthor
@DrEugeniaCheng
@DrMathochist
@DrMRFrancis
@drskyskull
@elakdawalla
@ErinPodolak
@evelynjlamb
@FlyingTrilobite
@HirokiSayama
@HowardBarnum
@Laelaps
@mandaYoho
@MarkCC
@michael_nielsen
@mikethemadbiol
@Myrmecos
@nervous_jessica
@phylogenomics
@planet4589
@pleunipennings
@PsiWavefunction
@pzmyers
@RaquelHRibeiro
@rebeccawatson
@rf
@sc_k
@scicurious
@seelix
@shanley
@TomLevenson
@TRyanGregory
@vihartvihart

Eric Downes was my gateway to the people and the puzzles with whom and which this thesis began.

My housemates put up with me. I'm not entirely sure why I deserved it.

For my family, all my love.

Lieutenant Columbo, Avatar Korra and Daria Morgendorffer were always there for me in the small hours of the morning.

Bibliography

- [1] T.-N. Coates, “Continuing education,” *The Atlantic* (2014) n.p. <http://www.theatlantic.com/education/archive/2014/10/continuing-education/381082/>.
- [2] W. James, *Pragmatism: A New Name for Some Old Ways of Thinking*. Penguin (2000 reprint), 1907. https://en.wikisource.org/wiki/Pragmatism:_A_New_Name_for_Some_Old_Ways_of_Thinking.
- [3] B. Allen, B. C. Stacey, and Y. Bar-Yam, “An information-theoretic formalism for multiscale structure in complex systems,” [arXiv:1409.4708 \[cond-mat.stat-mech\]](https://arxiv.org/abs/1409.4708). Reviews and systematizes the complexity profile and related concepts.
- [4] Y. Bar-Yam, “A mathematical theory of strong emergence using multiscale variety,” *Complexity* **9** (2004) no. 6, 15–24. <http://necsi.edu/research/multiscale/>.
- [5] Y. Bar-Yam, “Multiscale complexity/entropy,” *Advances in Complex Systems* **7** (2004) 47–63. <http://necsi.edu/research/multiscale/>.
- [6] Y. Bar-Yam, “Multiscale variety in complex systems,” *Complexity* **9** (2004) no. 4, 37–45. <http://necsi.edu/research/multiscale/>.
- [7] S. Gheorghiu-Svirschevski and Y. Bar-Yam, “Multiscale analysis of information correlations in an infinite-range, ferromagnetic Ising system,” *Physical Review E* **70** (2004) 066115. <http://necsi.edu/research/multiscale/>.
- [8] R. Metzler and Y. Bar-Yam, “Multiscale complexity of correlated Gaussians,” *Physical Review E* **71** (2005) 046114. <http://necsi.edu/research/multiscale/>.
- [9] S. R. Eddy, “The C-value paradox, junk DNA and ENCODE,” *Current Biology* **22** (2012) no. 21, R898–99.
- [10] A. F. Palazzo and T. R. Gregory, “The case for junk DNA,” *PLOS Genetics* **10** (2014) no. 5, e1004351.
- [11] J. Ross-Ibarra, M. Tenaillon, and B. S. Gaut, “Historical divergence and gene flow in the genus *Zea*,” *Genetics* **181** (2009) no. 4, 1399–1413.
- [12] M. Tenaillon, M. B. Hufford, B. S. Gaut, and J. Ross-Ibarra, “Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*,” *Genome Biology and Evolution* **3** (2011) 219–29.
- [13] W. F. Doolittle, “Is junk DNA bunk? A critique of ENCODE,” *Proceedings of the National Academy of Sciences* **110** (2013) no. 14, 5294–5300, [PMC:3619371](https://pubmed.ncbi.nlm.nih.gov/23619371/).
- [14] D. Graur *et al.*, “On the immortality of television sets: “Function” in the human genome according to the evolution-free gospel of ENCODE,” *Genome Biology and Evolution* **5** (2013) no. 3, 578–90.
- [15] L. D. Hurst, “Open questions: A logic (or lack thereof) of genome organization,” *BMC Biology* **11** (2013) 58.

Bibliography

- [16] D.-K. Niu and L. Jiang, “Can ENCODE tell us how much junk DNA we carry in our genome?,” *Biochemical and Biophysical Research Communications* **430** (2013) no. 4, 1340–43.
- [17] A. F. Palazzo and E. S. Lee, “Non-coding RNA: what is functional and what is junk?,” *Frontiers in Genetics* **6** (2015) no. 2, epub ahead of print.
- [18] D. Graur, Y. Zheng, and R. B. R. Azevedo, “An evolutionary classification of genome function,” *Genome Biology and Evolution* **7** (2015) epub ahead of print.
- [19] T. A. Elliott and T. R. Gregory, “What’s in a genome? the C-value enigma and the evolution of eukaryotic genome content,” *Philosophical Transactions B* **370** (2015) 20140331.
- [20] S. K. Sessions and A. Larson, “Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution,” *Evolution* **41** (1987) no. 6, 1239–51.
- [21] L. Stryer, *Biochemistry*. W. H. Freeman, third ed., 1988.
- [22] G. Sella and D. H. Ardell, “The coevolution of genes and genetic codes: Crick’s frozen accident revisited,” *Journal of Molecular Evolution* **63** (2006) no. 3, 297–313.
- [23] H. H. Guo, J. Choe, and L. A. Loeb, “Protein tolerance to random amino acid change,” *Proceedings of the National Academy of Sciences* **101** (2004) no. 25, 9205–10, [PMC:438954](#).
- [24] J. Maynard Smith, “Natural selection and the concept of a protein space,” *Nature* **225** (1970) 563–64.
- [25] W. K. Wootters, “Quantum measurements and finite geometry,” *Foundations of Physics* **36** (2006) no. 1, 112–26, [arXiv:quant-ph/0406032](#).
- [26] J. Ueberberg, *Foundations of Incidence Geometry: Projective and Polar Spaces*. Springer, 2010.
- [27] T. Bar-on, “Discrete Wigner function by symmetric informationally complete positive operator valued measure,” *Journal of Mathematical Physics* **50** (2009) no. 7, 072106.
- [28] D. J. A. Welsh, *Matroid Theory*. Academic Press, 1976.
- [29] S. Pruett-Jones and M. Pruett-Jones, “Sexual competition and courtship disruptions: why do male bowerbirds destroy each other’s bowers?,” *Animal Behaviour* **47** (1994) no. 3, 607–20.
- [30] L. Gonick, “The bowerbird’s dilemma,” *Discover* **1994** (1994) no. 10, 138–9. <http://www.msri.org/ext/larryg/pages/15.htm>.
- [31] D. Harmon, M. De Aguiar, D. Chinellato, D. Braha, I. Epstein, and Y. Bar-Yam, “Predicting economic market crises using measures of collective panic,” *PLOS One* **10** (2015) no. 7, e0131871, [arXiv:1102.2620 \[q-fin.ST\]](#). <http://www.necsi.edu/research/economics/economicpanic.html>.
- [32] M. A. M. de Aguiar and Y. Bar-Yam, “The Moran model as a dynamical process on networks and its implications for neutral speciation,” *Physical Review E* **84** (2011) 031901, [arXiv:1012.3913 \[q-bio.PE\]](#).
- [33] B. Allen and C. E. Tarnita, “Measures of success in a class of evolutionary models with fixed population size and structure,” *Journal of Mathematical Biology* **68** (2014) no. 1–2, 109–43.

- <http://scholar.princeton.edu/ctarnita/files/AxiomSuccessPublished.pdf>.
This paper builds up in a nice way the treatment of evolution as a stochastic process. The point made near the end about the way the Price equation is often misleadingly written is an important one.
- [34] P. E. Turner and L. Chao, “Prisoner’s dilemma in an RNA virus,” *Nature* **398** (1999) 441–43.
 - [35] D. Grieg and M. Travisano, “The prisoner’s dilemma and polymorphism in yeast *SUC* genes,” *Proceedings of the Royal Society B* **271** (2004) S25–S26, [PMC:1810003](#).
 - [36] W. R. Ashby, *An Introduction to Cybernetics*. Chapman & Hall, London, UK, 1956. <http://pespmc1.vub.ac.be/ASHBBOOK.html>. Insightful, speculative science from decades past has an engaging zeerust all its own.
 - [37] B. Allen and M. A. Nowak, “Games on graphs,” *EMS Surveys in Mathematical Sciences* **1** (2014) no. 1, 113–51.
 - [38] W. R. Tschinkel and E. O. Wilson, “Scientific natural history: Telling the epics of nature,” *BioScience* **64** (2014) no. 5, 438–43.
 - [39] B. C. Stacey and Y. Bar-Yam, “Principles of security: Human, cyber, and biological,” Tech. Rep. 2008-06-01, NECSI, 2008. [1303.2682 \[cs.CR\]](#). <http://www.necsi.edu/research/military/cyber/>.
 - [40] Y. Bar-Yam, “Complexity of military conflict: Multiscale complex systems analysis of littoral warfare,” tech. rep., NECSI, 2003. http://www.necsi.edu/projects/yaneer/SSG_NECSI_3_Litt.pdf.
 - [41] S. Levin, “The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture,” *Ecology* **73** (1992) no. 6, 1943–67.
 - [42] U. Dieckmann and R. Law, “Relaxation projections and the method of moments,” in Dieckmann *et al.* [108], pp. 412–55.
 - [43] S. Wright, “*Tempo and Mode in Evolution: A Critical Review*,” *Ecology* **26** (1945) no. 4, 415–19, [JSTOR:1931666](#).
 - [44] D. L. Hartl and A. G. Clark, *Principles of Population Genetics*. Sinauer Associates, 2007.
 - [45] K. M. Page and M. Nowak, “Unifying Evolutionary Dynamics,” *Journal of Theoretical Biology* **219** (2002) 93–8, [PMID:12392978](#).
 - [46] P. Bijma and M. Wade, “The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection,” *Journal of Evolutionary Biology* **21** (2008) 1175–88, [PMID:18547354](#). In the context of MLS-A scenarios, and considering only generation-to-generation changes, ideas which sound very different when said in words are mathematically interconvertible, with only linear transformations of coordinates. Recommended with caveats about the distinction between MLS-A and MLS-B, the terminology of “relatedness” versus “assortment” and the solecism of writing the Price equation with covariances.
 - [47] J. A. Damore and J. Gore, “Understanding microbial cooperation,” *Journal of Theoretical Biology* **299** (2012) 31–41, [PMID:21419783](#). <http://gorelab.homestead.com/Papers/UnderstandingMicrobialCooperation.pdf>. Recommended with the proviso that writing the Price equation using covariance notation, though common, can be misleading. This is mentioned in the text, but it deserves special emphasis.

Bibliography

- [48] Y. Bar-Yam, “Formalizing the gene-centered view of evolution,” *Advances in Complex Systems* **2** (1999) 277–81, [arXiv:physics/0002016](https://arxiv.org/abs/physics/0002016).
<http://necsi.edu/research/evoeco/>.
- [49] M. Kardar, *Statistical Physics of Particles*. Cambridge University Press, 2007.
<http://www.mit.edu/~kardar/teaching/index.html>. The first volume of two in a textbook series.
- [50] C. Goodnight *et al.*, “Evolution in spatial predator–prey models and the “prudent predator”: the inadequacy of steady-state organism fitness and the concept of individual and group selection,” *Complexity* **13** (2008) 23–44.
<http://necsi.edu/research/evoeco/>.
- [51] S. Lion and M. van Baalen, “Self-structuring in spatial evolutionary ecology,” *Ecology Letters* **11** (2008) 277–95, PMID:18070102.
- [52] O. Givan *et al.*, “Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches,” *Journal of Theoretical Biology* **288** (2011) 21–28, PMID:21840323.
- [53] G. C. Williams, *Natural Selection: Domains, Levels and Challenges*. Oxford University Press, 1992.
- [54] H. Sayama *et al.*, “Symmetry breaking and coarsening in spatially distributed evolutionary processes including sexual reproduction and disruptive selection,” *Physical Review E* **62** (2000) 7065–69.
- [55] J. D. Van Dyken, T. A. Linksvayer, and M. J. Wade, “Kin selection–mutation balance: A model for the origin, maintenance and consequences of social cheating,” *American Naturalist* **177** (2011) no. 3, 288–300, PMID:21460538.
- [56] M. Archetti and I. Scheuring, “Coexistence of cooperation and defection in public goods games,” *Evolution* **65** (2011) no. 4, 1140–48, PMID:21062277.
- [57] P. D. Taylor and S. A. Frank, “How to Make a Kin Selection Model,” *Journal of Theoretical Biology* **180** (1996) 27–37, PMID:8763356.
- [58] H. Kokko, K. U. Heubel, and D. J. Rankin, “How populations persist when asexuality requires sex: the spatial dynamics of coping with sperm parasites,” *Proceedings of the Royal Society B* **275** (2008) 817–25, PMID:18182369.
- [59] G. Wild, A. Gardner, and S. A. West, “Adaptation and the evolution of parasite virulence in a connected world,” *Nature* **459** (2009) 983–6, PMID:19474791.
- [60] R. Levins, “Some demographic and genetic consequences of environmental heterogeneity for biological control,” *Bulletin of the Entomological Society of America* **15** (1969) 237.
- [61] A. Platt *et al.*, “The scale of population structure in *Arabidopsis thaliana*,” *PLOS Genetics* **6** (2010) no. 2, e1000843.
- [62] J. M. Halley *et al.*, “Uses and abuses of fractal methodology in ecology,” *Ecology Letters* **7** (2004) 254–71.
- [63] T. M. Scanlon, K. K. Caylor, S. A. Levin, and I. Rodriguez-Iturbe, “Positive feedbacks promote power-law clustering of Kalahari vegetation,” *Nature* **449** (2007) no. 7159, 209–12, PMID:17851523.

- [64] C. Reigada and M. A. M. de Aguiar, “Host-parasitoid persistence over variable spatio-temporally susceptible habitats: bottom-up effects of ephemeral resources,” *Oikos* **121** (2012) 1665–79.
<http://www.ifi.unicamp.br/~aguiar/Abstracts/2011h.html>.
- [65] G. Ichinose, M. Saito, H. Sayama, and D. S. Wilson, “Adaptive long-range migration promotes cooperation under tempting conditions,” *Scientific Reports* **3** (2013) 2509, [arXiv:1306.0072](https://arxiv.org/abs/1306.0072) [q-bio.PE].
- [66] A. Escalante *et al.*, “The dynamics of mutations associated with anti-malarial drug resistance in *Plasmodium falciparum*,” *Trends in Parasitology* **25** (2009) no. 12, 557–63, [PMC:2881657](https://pubmed.ncbi.nlm.nih.gov/2881657/).
- [67] T. Gross and B. Blasius, “Adaptive coevolutionary networks: a review,” *Journal of the Royal Society Interface* **5** (2008) no. 20, 259–71, [PMC:2405905](https://pubmed.ncbi.nlm.nih.gov/2405905/).
- [68] T. Gross and H. Sayama, eds., *Adaptive Networks: Theory, Models, Applications*. Springer, 2009.
- [69] O. Gräser, C. Xu, and P. M. Hui, “Analytic approach to co-evolving dynamics in complex networks: dissatisfied adaptive snowdrift game,” *New Journal of Physics* **13** (2011) no. 8, 083015.
- [70] G. Demirel, F. Vazquez, G. A. Böhme, and T. Gross, “Moment-closure approximations for discrete adaptive networks,” [arXiv:1211.0449](https://arxiv.org/abs/1211.0449) [nlin.AO]. More elaborate moment closures can actually perform worse, and even when approximations are satisfactory, it might not be possible to find one approximation which works across the whole parameter space.
- [71] H. Sayama, *Introduction to the Modeling and Analysis of Complex Systems*. Open SUNY, 2015. <http://textbooks.opensuny.org/introduction-to-the-modeling-and-analysis-of-complex-systems/>.
- [72] J. Werfel and Y. Bar-Yam, “The evolution of reproductive restraint through social communication,” *PNAS* **101** (2004) no. 30, 11019–24, [PMID:15256603](https://pubmed.ncbi.nlm.nih.gov/15256603/).
http://www.necsi.edu/projects/evolecol/altruism_pnaspr.html.
- [73] A. Jiménez-Dalmaroni and H. Hinrichsen, “Epidemic processes with immunization,” *Physical Review E* **68** (2003) 036103, [arXiv:cond-mat/0304113](https://arxiv.org/abs/cond-mat/0304113).
- [74] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions, Volume 1: Absorbing Phase Transitions*. Springer, 2008.
- [75] H. Kokko and K. U. Heubel, “Prudent males, group adaptation, and the tragedy of the commons,” *Oikos* **120** (2011) 641–56.
- [76] Y. Haraguchi and A. Sasaki, “The Evolution of Parasite Virulence and Transmission Rate in a Spatially Structured Population,” *Journal of Theoretical Biology* **203** (2000) 85–96, [PMID:10704294](https://pubmed.ncbi.nlm.nih.gov/10704294/).
- [77] H. Sayama, M. A. M. de Aguiar, Y. Bar-Yam, and M. Baranger, “Spontaneous pattern formation and genetic invasion in locally mating and competing populations,” *Physical Review E* **65** (2002) 051919.
- [78] M. A. M. de Aguiar, M. Baranger, Y. Bar-Yam, and H. Sayama, “Robustness of spontaneous pattern formation in spatially distributed genetic populations,” *Brazilian Journal of Physics* **33** (2003) no. 3, 514–20.

Bibliography

- [79] H. Sayama, L. Kaufman, and Y. Bar-Yam, “Spontaneous pattern formation and genetic diversity in habitats with irregular geographical features,” *Conservation Biology* **17** (2003) 893–900.
- [80] C. G. Jones, J. H. Lawton, and M. Shachak, “Organisms as ecosystem engineers,” *Oikos* **69** (1994) no. 3, 373–86, [JSTOR:3545850](#).
- [81] D. L. Strayer, V. T. Eviner, J. M. Jeschke, and M. L. Pace, “Understanding the long-term effects of species invasions,” *Trends in Ecology and Evolution* **21** (2006) no. 11, 645–51, [PMID:16859805](#).
- [82] D. M. Post and E. P. Palkovacs, “Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play,” *Philosophical Transactions of the Royal Society B* **364** (2009) 1629–40.
- [83] R. M. Pringle *et al.*, “Spatial pattern enhances ecosystem functioning in an African savanna,” *PLOS Biology* **8** (2010) no. 5, e1000377.
- [84] B. Allen and M. A. Nowak, “Cooperation and the fate of microbial societies,” *PLOS Biology* **11** (2013) no. 4, e1001549. This is an introductory overview of what can happen when ecological and evolutionary processes occur on comparable timescales and feed back upon one another. It summarizes recent experimental and model-building work on the topic, as realised in *Saccharomyces cerevisiae*.
- [85] B. J. Callahan, T. Fukami, and D. S. Fisher, “Rapid evolution of adaptive niche construction in experimental microbial populations,” *Evolution* **68** (2014) no. 11, 3307–16.
- [86] J. T. Cronin and J. D. Reeve, “Host-parasitoid spatial ecology: a plea for a landscape-level synthesis,” *Proceedings of the Royal Society B* **272** (2005) no. 1578, 2225–35, [PMID:16191634](#).
- [87] B. Kerr, C. Neuhauser, B. J. M. Bohannan, and A. M. Dean, “Local migration promotes competitive restraint in a host–pathogen ‘tragedy of the commons’,” *Nature* **442** (2006) 75–8, [PMID:16823452](#).
- [88] S. M. Messinger and A. Ostling, “The consequences of spatial structure for the evolution of pathogen transmission rate and virulence,” *American Naturalist* **174** (2009) no. 4, 441–54.
- [89] S. Lion and M. Boots, “Are parasites “prudent” in space?,” *Ecology Letters* **13** (2010) 1245–55.
- [90] S. M. Messinger and A. Ostling, “The influence of host demography, pathogen virulence, and relationships with pathogen virulence on the evolution of pathogen transmission in a spatial context,” *Evolutionary Ecology* **27** (2012) no. 2, 353–80.
- [91] E. Rauch and Y. Bar-Yam, “Long-range interactions and evolutionary stability in a predator–prey system,” *Physical Review E* **73** (2006) 020903, [PMID:16605322](#). <http://necsi.edu/research/evoeco/>.
- [92] S. Heilmann, K. Sneppen, and S. Krishna, “Sustainability of virulence in a phage-bacteria ecosystem,” *Journal of Virology* **84** (2010) no. 6, 3016–22.
- [93] E. Rauch, H. Sayama, and Y. Bar-Yam, “Relationship between Measures of Fitness and Time Scale in Evolution,” *Physical Review Letters* **88** (2002) 228101, [PMID:12059453](#).

- [94] E. Rauch, H. Sayama, and Y. Bar-Yam, “Dynamics and Genealogy of Strains in Spatially Extended Host–Pathogen Models,” *Journal of Theoretical Biology* **221** (2003) 655–64, PMID:12713947.
- [95] M. A. M. de Aguiar, E. Rauch, and Y. Bar-Yam, “Invasion and Extinction in the Mean Field Approximation for a Spatial Host–Pathogen Model,” *Journal of Statistical Physics* **114** (2004) no. 5/6, 1417–51.
- [96] R. Froissart, J. Doumayrou, F. Vuillaume, S. Alizon, and Y. Michalakakis, “The virulence–transmission trade-off in vector-borne plant viruses: a review of (non-)existing studies,” *Philosophical Transactions of the Royal Society B* **365** (2010) no. 1548, 1907–18, PMC:2880117.
- [97] M. K. Asplen *et al.*, “Do trade-offs have explanatory power for the evolution of organismal interactions?,” *Evolution* **66** (2012) no. 5, 1297–1307.
- [98] D. M. Hawley *et al.*, “Parallel patterns of increased virulence in a recently emerged wildlife pathogen,” *PLOS Biology* **11** (2013) no. 5, e1001570.
- [99] M. Mobilia, I. T. Georgiev, and U. C. Täuber, “Phase transitions and spatio-temporal fluctuations in stochastic lattice Lotka–Volterra models,” *Journal of Statistical Physics* **128** (2006) no. 1–2, 447–83, arXiv:q-bio/0512039.
- [100] M. A. M. de Aguiar, E. M. Rauch, and Y. Bar-Yam, “Mean-field approximation to a spatial host-pathogen model,” *Physical Review E* **67** (2003) no. 4, 047102, arXiv:1307.5335 [q-bio.PE]. <http://necsi.edu/research/evoeco/>.
- [101] M. A. M. de Aguiar, E. Rauch, B. C. Stacey, and Y. Bar-Yam, “Erratum: Mean-field approximation to a spatial host-pathogen model,” *Physical Review E* **88** (2013) no. 3, 039901(E), arXiv:1307.5335 [q-bio.PE]. Finding the problem, figuring out the correct solution and writing the erratum: 2 days. Time it takes *Physical Review E* to publish that erratum: 2 months. Also, they took a great deal of convincing before they agreed that the person who found the problem and wrote the corrections should be listed as an author. Their copy editors complained that $\bar{\omega}$ was never used in the original paper (it was), and that $\bar{\chi}$ is never defined (it was).
- [102] E. Arashiro and T. Tomé, “Threshold of coexistence and critical behavior of a predator-prey cellular automaton,” *Journal of Physics A* **40** (2007) 887–900, arXiv:cond-mat/0607360.
- [103] M. van Baalen and D. A. Rand, “The unit of selection in viscous populations and the evolution of altruism,” *Journal of Theoretical Biology* **193** (1998) 631–48, PMID:9750181.
- [104] M. van Baalen, “Pair Approximations for Different Spatial Geometries,” in Dieckmann *et al.* [108], pp. 359–85.
- [105] G. Rozhnova and A. Nunes, “Population dynamics on random networks: simulations and analytical models,” *European Physical Journal B* **74** (2010) no. 2, 235–42, arXiv:0907.0335 [q-bio]. An example of a case where the results of pair approximation are qualitatively incorrect.
- [106] B. Allen, *Studies in the Mathematics of Evolution and Biodiversity*. PhD thesis, Boston University, 2010. <http://proquest.umi.com/pqdweb?did=2071736811&Fmt=2&clientId=5482&RQT=309>.

Bibliography

- [107] S. B. L. Araujo, G. M. Viswanathan, and M. A. M. de Aguiar, “Home range evolution and its implication in population outbreaks,” *Philosophical Transactions of the Royal Society A* **368** (2010) 5661–77. Pair approximations done carefully, with acknowledgements of their limitations, too.
- [108] U. Dieckmann, R. Law, and J. A. J. Metz, eds., *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge University Press, 2000.
- [109] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, “Statistical mechanics of population,” *Progress of Theoretical Physics* **88** (1992) no. 6, 1035–49.
- [110] A.-L. Do and T. Gross, “Contact processes and moment closure on adaptive networks,” in Gross and Sayama [68], pp. 191–208.
- [111] M. A. Buice and J. D. Cowan, “Statistical mechanics of the neocortex,” *Progress in Biophysics and Molecular Biology* **99** (2009) 53–86, PMID:19695282.
- [112] P. J. Dodd and N. M. Ferguson, “A many-body field theory approach to stochastic models in population biology,” *PLOS ONE* **4** (2009) no. 9, e6855, PMID:19730742.
- [113] G. Szabó and G. Fáth, “Evolutionary games on graphs,” *Physics Reports* **446** (2007) 97–216, [arXiv:cond-mat/0607344](https://arxiv.org/abs/cond-mat/0607344).
- [114] B. Allen, A. Traulsen, C. E. Tarnita, and M. A. Nowak, “How mutation affects evolutionary games on graphs,” *Journal of Theoretical Biology* **299** (2012) 97–105, PMID:21473871.
<http://www.people.fas.harvard.edu/~ballen/GraphMutInPress.pdf>.
- [115] P. E. Smaldino, J. C. Schank, and R. McElreath, “Increased costs of cooperation help cooperators in the long run,” *American Naturalist* **181** (2013) no. 4, 451–63.
<http://xcelab.net/rmpubs/Smaldino%20Am%20Nat%202013.pdf>. One of many papers which points out the shortcomings of “moment closure” techniques, this also makes a key point about fitness being a question of timescale.
- [116] H. J. Poethke, B. Pfenning, and T. Hovestadt, “The relative contribution of individual and kin selection to the evolution of density-dependent dispersal rates,” *Evolutionary Ecology Research* **9** (2007) 41–50, OPUS:2010/4822.
- [117] S. Lion, “Relatedness in spatially structured populations with empty sites: An approach based on spatial moment equations,” *Journal of Theoretical Biology* **260** (2009) 121–31.
- [118] S. Davis, P. Trapman, H. Leirs, M. Begon, and J. A. P. Heesterbeek, “The abundance threshold for plague as a critical percolation phenomenon,” *Nature* **454** (2008) 634–37.
- [119] D. J. Salkeld, M. Salathé, P. Stapp, and J. H. Jones, “Plague outbreaks in prairie dog populations explained by percolation thresholds of alternate host abundance,” *Proceedings of the National Academy of Sciences* **107** (2010) no. 32, 14247–50.
- [120] F. M. Neri *et al.*, “The effect of heterogeneity on invasion in spatial epidemics: from theory to evidence in a model system,” *PLOS Computational Biology* **7** (2011) no. 9, e1002174.
- [121] H. Hinrichsen, “Nonequilibrium critical phenomena and phase transitions into absorbing states,” *Advances in Physics* **49** (2000) no. 7, 815–958, [arXiv:cond-mat/0001070](https://arxiv.org/abs/cond-mat/0001070).
- [122] M. A. Saif and P. M. Gade, “Dynamic phase transition in the prisoner’s dilemma on a lattice with stochastic modifications,” *Journal of Statistical Mechanics* (2010) P03016, [arXiv:0910.0955](https://arxiv.org/abs/0910.0955) [cond-mat.stat-mech].

- [123] J. Wendykier, A. Lipowski, and A. L. Ferreira, “Coexistence and critical behaviour in a lattice model of competing species,” *Physical Review E* **83** (2011) no. 3, 031904, [arXiv:1010.2538 \[q-bio.PE\]](#).
- [124] A. Szolnoki, G. Szabó, and M. Perc, “Phase diagrams for the spatial public good game with pool-punishment,” *Physical Review E* **83** (2011) no. 3, 036101, [arXiv:1102.0624 \[physics.soc-ph\]](#).
- [125] P. Grassberger, H. Chaté, and G. Rousseau, “Spreading in media with long-time memory,” *Physical Review E* **55** (1997) no. 3, 2488–95.
- [126] H. Kesten, “The critical probability of bond percolation on the square lattice equals $1/2$,” *Communications in Mathematical Physics* **74** (1980) no. 1, 41–59.
- [127] G. Grimmett, *Percolation*. Springer, 1999.
- [128] M. E. J. Newman, “Spread of epidemic disease on networks,” *Physical Review E* **66** (2002) 016128, [arXiv:cond-mat/0205009](#).
- [129] J. P. Gleeson, “High-accuracy approximation of binary-state dynamics on networks,” *Physical Review Letters* **107** (2011) 068701, [arXiv:1104.1537 \[physics.soc-ph\]](#).
- [130] M. Boguñá, C. Castellano, and R. Pastor-Satorras, “Nature of the epidemic threshold for the Susceptible-Infected-Susceptible dynamics in networks,” *Physical Review Letters* **111** (2013) 068701, [arXiv:1305.4819 \[physics.soc-ph\]](#).
- [131] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, “Heterogeneous pair-approximation for the contact process on complex networks,” *New Journal of Physics* **16** (2014) 053006.
- [132] C. A. Lugo and A. J. McKane, “Quasicycles in a spatial predator-prey model,” *Physical Review E* **78** (2008) 051911, [arXiv:0806.1287 \[q-bio.PE\]](#).
- [133] T. Biancalani, D. Fanelli, and F. D. Patti, “Stochastic Turing patterns in the Brusselator model,” *Physical Review E* **81** (2010) 046215, [arXiv:0910.4984 \[cond-mat.stat-mech\]](#).
- [134] T. Butler and N. Goldenfeld, “Fluctuation-driven Turing patterns,” *Physical Review E* **84** (2011) 011112, [arXiv:1011.0466 \[q-bio.OT\]](#). Equations (50) and (52) in the journal version are incorrect.
- [135] M. Pascual, M. Roy, F. Guichard, and G. Flierl, “Cluster size distributions: signatures of self-organization in spatial ecologies,” *Philosophical Transactions of the Royal Society B* **357** (2002) 657–66.
- [136] J. Alstott, E. Bullmore, and D. Plenz, “powerlaw: a Python package for analysis of heavy-tailed distributions,” *PLOS One* **9** (2014) no. 1, e85777, [arXiv:1305.0215 \[physics.data-an\]](#).
- [137] S. Gueron and S. A. Levin, “The dynamics of group formation,” *Mathematical Biosciences* **128** (1995) no. 1–2, 243–64.
- [138] Q. Ma, A. Johansson, and D. J. T. Sumpter, “A first principles derivation of animal group size distributions,” *Journal of Theoretical Biology* **283** (2011) 35–43.
- [139] M. Visser, “Zipf’s law, power laws and maximum entropy,” *New Journal of Physics* **15** (2013) 043021.
- [140] D. S. Wilson and R. C. Lewontin, “The Spandrels of San Marco revisited: An interview with Richard C. Lewontin,” *This View of Life* (2015) n.p. <https://evolution-institute.org/article/the-spandrels-of-san-marco-revisited-an-interview-with-richard-c-lewontin/>.

Bibliography

- [141] H. A. Orr, “Fitness and its role in evolutionary genetics,” *Nature Reviews Genetics* **10** (2009) 531–39.
- [142] J. Smith, J. D. Van Dyken, and P. C. Zee, “A generalization of Hamilton’s rule for the evolution of microbial cooperation,” *Science* **328** (2010) 1700–03.
- [143] C. A. Fogle, J. L. Nagle, and M. M. Desai, “Clonal interference, multiple mutations and adaptation in large asexual populations,” *Genetics* **180** (2008) no. 4, 2163–73, PMID:18832359.
- [144] R. Maddamsetti, R. E. Lenski, and J. E. Barrick, “Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*,” *Genetics* (2015) epub ahead of print.
- [145] K. Hammerschmidt, C. J. Rose, B. Kerr, and P. B. Rainey, “Life cycles, fitness decoupling and the evolution of multicellularity,” *Nature* **515** (2014) 75–79.
- [146] J. Fox, “Recent ecology & evolution papers I thought were really good,” *Dynamic Ecology* (2015) n.p. <https://dynamicecology.wordpress.com/2015/03/16/recent-papers-i-thought-were-really-good/>.
- [147] A. Grafen, “A geometric view of relatedness,” *Oxford Surveys in Evolutionary Biology* **2** (1985) 28–89. <http://users.ox.ac.uk/~grafen/cv/oseb.pdf>.
- [148] B. Simon, J. A. Fletcher, and M. Doebeli, “Hamilton’s rule in multi-level selection models,” *Journal of Theoretical Biology* **299** (2012) 55–63, PMID:21820447. Builds quantitative models of an MLS-B scenario: there are explicit group-level dynamics, but the set of groups is unstructured.
- [149] H. K. Reeve and B. Hölldobler, “The emergence of a superorganism through intergroup competition,” *PNAS* **104** (2007) no. 23, 9736–40, PMID:1887545.
- [150] B. Allen, M. A. Nowak, and E. O. Wilson, “Limitations of inclusive fitness,” *Proceedings of the National Academy of Sciences* **110** (2013) no. 50, 20135–39. You can fit a straight line to any cloud of data points, but why should you want to?
- [151] B. Allen, “What’s the deal with inclusive fitness theory?,” *Plektix* (2013) n.p. <http://plektix.fieldofscience.com/2013/12/whats-deal-with-inclusive-fitness-theory.html>.
- [152] M. Wade *et al.*, “Multilevel and kin selection in a connected world,” *Nature* **463** (2010) E8. <http://necsi.edu/research/evoeco/>.
- [153] B. Simon, J. A. Fletcher, and M. Doebeli, “Towards a general theory of group selection,” *Evolution* **67** (2013) no. 6, 1561–72. <http://math.ucdenver.edu/~bsimon/papers.html>.
- [154] A. Nielsen and B. Simon, “Numerical solutions and animations of group selection dynamics,” *Evolutionary Ecology Research* **44** (2012) 757–68. <http://math.ucdenver.edu/~bsimon/papers.html>.
- [155] D. Campobello, J. F. Hare, and M. Sarà, “Social phenotype extended to communities: Expanded multilevel social selection analysis reveals fitness consequences of interspecific interactions,” *Evolution* **69** (2015) no. 4, 916–25.
- [156] G. Wild and P. D. Taylor, “Fitness and evolutionary stability in game theoretic models of finite populations,” *Proceedings of the Royal Society of London B* **271** (2004) 2345–9, PMID:15590589.

- [157] S. D. Mylius and O. Diekmann, “The resident strikes back: Invader-induced switching of resident attractor,” *Journal of Theoretical Biology* **211** (2001) 297–311.
<http://sido.mylius.nl/downloads/rstrbftp.pdf>.
- [158] S. A. H. Geritz, M. Gyllenberg, F. J. A. Jacobs, and K. Parvinen, “Invasion dynamics and attractor inheritance,” *Journal of Mathematical Biology* **44** (2002) 548–60.
- [159] S. A. H. Geritz, “Resident-invader dynamics and the coexistence of similar strategies,” *Journal of Mathematical Biology* **50** (2005) 67–82.
- [160] M. Doebeli, “Invasion of rare mutants does not imply their evolutionary success: a counterexample from metapopulation theory,” *Journal of Evolutionary Biology* **11** (1997) no. 3, 389–401. <http://www.math.ubc.ca/~doebeli/reprints/Doe26.pdf>.
- [161] S. Lessard, “Long-term stability from fixation probabilities in finite populations: New perspectives for ESS theory,” *Theoretical Population Biology* **68** (2005) 19–27, PMID:16023912.
- [162] T. Antal and I. Scheuring, “Fixation of strategies in an evolutionary game in finite populations,” *Bulletin of Mathematical Biology* **68** (2006) no. 8, 1923–44, [arXiv:q-bio/0509008](http://arxiv.org/abs/q-bio/0509008).
- [163] C. J. Paley, S. N. Taraskin, and S. R. Elliott, “Temporal and dimensional effects in evolutionary graph theory,” *Physical Review Letters* **98** (2007) 098103, [arXiv:q-bio/0604009](http://arxiv.org/abs/q-bio/0604009).
- [164] J. M. Ponciano, H. J. La, P. Joyce, and L. J. Forney, “Evolution of diversity in spatially structured *Escherichia coli* populations,” *Applied and Environmental Microbiology* **75** (2009) no. 19, 6047–54, PMID:19648364.
- [165] M. Raghib, N. A. Hill, and U. Diekmann, “Multiscale maximum entropy closure for locally regulated space-time point process models of plant population dynamics,” *Journal of Mathematical Biology* **62** (2010) no. 5, 605–53, PMID:20446087.
- [166] H.-K. Janssen and U. C. Täuber, “The field theory approach to percolation processes,” *Annals of Physics* **315** (2005) 147–92, [arXiv:cond-mat/0409670](http://arxiv.org/abs/cond-mat/0409670).
- [167] T. Gross, C. J. D. D’Lima, and B. Blasius, “Epidemic dynamics on an adaptive network,” *Physical Review Letters* **96** (2006) no. 20, 208701, [arXiv:q-bio/0512037](http://arxiv.org/abs/q-bio/0512037).
- [168] L. B. Shaw and I. B. Schwartz, “Noise induced dynamics in adaptive networks with applications to epidemiology,” in Gross and Sayama [68], pp. 209–27.
- [169] L. B. Shaw and I. B. Schwartz, “Enhanced vaccine control of epidemics in adaptive networks,” *Physical Review E* **81** (2010) no. 4, 046120, PMC:2931598.
- [170] C. Kamp, “Untangling the interplay between epidemic spread and transmission network dynamics,” *PLOS Computational Biology* **6** (2010) no. 11, e1000984, PMC:2987842.
- [171] S. Van Segbroeck, F. C. Santos, and J. M. Pacheco, “Adaptive contact networks change effective disease infectiousness and dynamics,” *PLOS Computational Biology* **6** (2010) no. 8, e1000895, PMC:2924249.
- [172] B. Wu *et al.*, “Evolution of cooperation on stochastic dynamical networks,” *PLOS ONE* **5** (2010) no. 6, e11187, PMC:2894855.
- [173] K. Feh1, D. J. van der Post, and D. Semmann, “Co-evolution of behaviour and social network structure promotes human cooperation,” *Ecology Letters* **14** (2011) 546–51, PMID:21463459.

Bibliography

- [174] A. Traulsen, F. C. Santos, and J. M. Pacheco, “Evolutionary games in self-organizing populations,” in Gross and Sayama [68], pp. 253–68.
- [175] P. Holme and G. Ghoshal, “The diplomat’s dilemma: Maximal power for minimal effort in social networks,” in Gross and Sayama [68], pp. 269–88.
- [176] R. A. Fisher, *The Genetical Theory of Natural Selection*. Clarendon, 1930.
- [177] C. Ainsworth, “Sex redefined,” *Nature* **518** (2015) no. 7539, 288–91.
- [178] J. B. Xavier and K. R. Foster, “Cooperation and conflict in microbial biofilms,” *PNAS* **104** (2007) no. 3, 876–81, PMID:17210916.
- [179] B. Gönci *et al.*, “Viral epidemics in a cell culture: Novel high resolution data and their interpretation by a percolation theory based model,” *PLOS ONE* **5** (2010) no. 12, e15571, PMC:3004943.
- [180] D. Wodarz *et al.*, “Complex spatial dynamics of oncolytic viruses in vitro: Mathematical and experimental approaches,” *PLOS Computational Biology* **8** (2012) no. 6, e1002547.
- [181] M. A. M. de Aguiar, M. Baranger, E. M. Baptestini, L. Kaufman, and Y. Bar-Yam, “Global patterns of speciation and diversity,” *Nature* **460** (2009) 384–87.
<http://www.necsi.edu/research/evoeco/globalpatternsofspeci/index.html>.
- [182] A. B. Martins, M. A. M. de Aguiar, and Y. Bar-Yam, “Evolution and stability of ring species,” *PNAS* **110** (2013) no. 13, 5080–84.
<http://www.necsi.edu/research/evoeco/ringspecies/>.
- [183] J. D. Van Dyken, M. J. I. Müller, K. M. L. Mack, and M. M. Desai, “Spatial population expansion promotes the evolution of cooperation in an experimental prisoner’s dilemma,” *Current Biology* **23** (2013) no. 10, 919–23, arXiv:1311.2646 [q-bio.PE].
- [184] E. M. Baptestini, M. A. M. de Aguiar, and Y. Bar-Yam, “Conditions for neutral speciation via isolation by distance,” *Journal of Theoretical Biology* **335** (2013) 51–6.
<http://necsi.edu/research/evoeco/>.
- [185] B. Allen, C. Sample, Y. Dementieva, R. C. Medeiros, C. Paoletti, and M. A. Nowak, “The molecular clock of neutral evolution can be accelerated or slowed by asymmetric spatial structure,” *PLOS Computational Biology* **11** (2015) no. 2, e1004108.
- [186] J. Werfel, D. E. Ingber, and Y. Bar-Yam, “Programmed death is favored by natural selection in spatial systems,” *Physical Review Letters* **114** (2015) 238103.
<http://www.necsi.edu/research/evoeco/programmeddeath.html>.
- [187] J. Werfel, D. E. Ingber, and Y. Bar-Yam, “Theory and associated phenomenology for intrinsic mortality arising from natural selection,” arXiv:1506.03893 [q-bio.PE].
- [188] B. Momeni, A. J. Waite, and W. Shou, “Spatial self-organization favors heterotypic cooperation over cheating,” *eLife* **2** (2013) e00960.
- [189] G. Lambert, S. Vyawahare, and R. H. Austin, “Bacteria and game theory: the rise and fall of cooperation in spatially heterogeneous environments,” *Interface Focus* **4** (2014) no. 4, n.p.
- [190] J. van Gestel, F. J. Weissing, O. P. Kuipers, and A. T. Kovács, “Density of founder cells affects spatial pattern formation and cooperation in *Bacillus subtilis* biofilms,” *The ISME Journal* **8** (2014) 2069–79.

- [191] U. Carlsson-Granér and P. H. Thrall, “Host resistance and pathogen infectivity in host populations with varying connectivity,” *Evolution* **69** (2015) no. 4, 926–38.
- [192] B. C. Stacey, A. Gros, and Y. Bar-Yam, “Eco-evolutionary feedback in host–pathogen spatial dynamics,” [arXiv:1110.3845](https://arxiv.org/abs/1110.3845) [[nlin.CG](https://arxiv.org/archive/nlin)]. <http://necsi.edu/research/evoeco/>. The material from this preprint has been incorporated into this thesis.
- [193] B. Skyrms and R. Pemantle, “A dynamic model of social network formation,” *Proceedings of the National Academy of Sciences* **97** (2000) no. 16, 9340–46, [arXiv:math/0404101](https://arxiv.org/abs/math/0404101) [[math.PR](https://arxiv.org/archive/math)].
- [194] J. M. Pacheco, F. C. Santos, M. O. Souza, and B. Skyrms, “Evolutionary dynamics of collective action in n -person stag hunt dilemmas,” *Proceedings of the Royal Society B* **276** (2009) no. 1655, 315–21, [PMC:2674356](https://pubmed.ncbi.nlm.nih.gov/2674356/).
- [195] M. van Veelen and M. A. Nowak, “Multi-player games on the cycle,” *Journal of Theoretical Biology* **292** (2012) 116–28.
- [196] R. Durrett and S. Levin, “The Importance of Being Discrete (and Spatial),” *Theoretical Population Biology* **46** (1994) 363–94.
- [197] B. Allen and M. A. Nowak, “Games among relatives revisited,” *Journal of Theoretical Biology* **378** (2015) 103–16.
- [198] M. Archetti and I. Scheuring, “Review: Game theory of public goods in one-shot social dilemmas without assortment,” *Journal of Theoretical Biology* **299** (2011) 9–20, [PMID:21723299](https://pubmed.ncbi.nlm.nih.gov/21723299/).
- [199] M. Kardar, *Statistical Physics of Fields*. Cambridge University Press, 2007. <http://www.mit.edu/~kardar/teaching/index.html>. Second in a two-part series of textbooks.
- [200] A. Chen, A. Sanchez, L. Dai, and J. Gore, “Dynamics of a producer-freeloader ecosystem on the brink of collapse,” *Nature Communications* **5** (2014) 3713, [PMC:4063714](https://pubmed.ncbi.nlm.nih.gov/24063714/). Experimental observation of critical slowing down in a microbial ecosystem.
- [201] M. A. Nowak, C. E. Tarnita, and E. O. Wilson, “The evolution of eusociality,” *Nature* **466** (2010) no. 7310, 1057–62, [PMC:3279739](https://pubmed.ncbi.nlm.nih.gov/203279739/). Everything in this paper of interest to this thesis is actually in the appendices, because scientific publishing is like that.
- [202] “Cubic symmetric graphs (The Foster Census),” Webpage, 2001. <http://mapleta.maths.uwa.edu.au/~gordon/remote/foster/>.
- [203] B. L. Granovsky and N. Madras, “The noisy voter model,” *Stochastic Processes and Their Applications* **55** (1995) no. 1, 23–43.
- [204] R. Jeffrey, *Subjective Probability: The Real Thing*. Cambridge University Press, 2004. <http://www.princeton.edu/~bayerway/>.
- [205] C. M. Caves, “Probabilities as betting odds and the Dutch book,” Internal report, 2005. <http://info.phys.unm.edu/~caves/reports/reports.html>. Somewhat complementary to the approach we take here.
- [206] C. A. Fuchs and R. Schack, “Bayesian conditioning, the reflection principle, and quantum decoherence,” [arXiv:1103.5950](https://arxiv.org/abs/1103.5950) [[quant-ph](https://arxiv.org/archive/quant)]. Although it is motivated by quantum physics, the particularly quantum considerations do not explicitly enter until near the end.

Bibliography

- [207] A. Hodges, *Alan Turing: The Enigma*. Princeton University Press, 1983.
<http://www.turing.org.uk>.
- [208] S. L. Zabell, “Alan Turing and the central limit theorem,” *American Mathematical Monthly* **102** (1995) no. 6, 483–94, [JSTOR:2974762](#).
- [209] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009. <http://pgm.stanford.edu/>.
- [210] P. W. Humphreys, “Notes on the origin of the term ‘Dutch book’,” Webpage, 2008.
<http://people.virginia.edu/~pwh2a/dutch%20book%20origins.doc>.
- [211] C. A. Fuchs, M. Schlosshauer, and B. C. Stacey, *My Struggles with the Block Universe*. Preprint, 2014. [arXiv:1405.2390 \[quant-ph\]](#). Possibly the record-holder for length of an arXiv posting.
- [212] T. Tao, *Topics in Random Matrix Theory*. American Mathematical Society, 2012.
<http://terrytao.wordpress.com/books/topics-in-random-matrix-theory/>.
- [213] J. H. Conway, *On Numbers and Games*. Academic Press, 1976. I’ve yet to find a real connection between the kind of mathematical games that Conway talks about and the type of games discussed in this thesis, but I’m still looking.
- [214] B. C. van Fraassen, “Belief and the Will,” *The Journal of Philosophy* **81** (1984) no. 5, 235–56, [JSTOR:2026388](#).
- [215] B. Skyrms, “Dynamic coherence and probability kinematics,” *Philosophy of Science* **54** (1987) no. 1, 1–20, [JSTOR:187470](#).
- [216] P. Diaconis and S. L. Zabell, “Updating subjective probability,” *Journal of the American Statistical Association* **77** (1982) no. 380, 822–30, [JSTOR:2287313](#).
- [217] P. Diaconis and S. L. Zabell, “Some alternatives to Bayes’s rule,” in *Proceedings of the Second University of California, Irvine, Conference on Political Economy*, pp. 25–38. 1986.
- [218] M. Harper, “The replicator equation as an inference dynamic,” [arXiv:0911.1763 \[math.DS\]](#). Explains the formal mapping between the Bayes rule and the replicator equation.
- [219] J. C. Baez, “Information geometry,” *Azimuth* (2012) n.p.
<http://math.ucr.edu/home/baez/information/>.
- [220] J. Moriarty, *A Treatise on the Binomial Theorem*. Meyer and Company, 1867.
- [221] F. Eggenberger and G. Pólya, “Über die Statistik verketteter Vorgänge,” *Zeitschrift für Angewandte Mathematik und Mechanik* **3** (1923) no. 4, 279–89.
- [222] A. Collecchio, C. Cotar, and M. LiCalzi, “On a preferential attachment and generalized Pólya’s urn model,” *The Annals of Applied Probability* **23** (2013) no. 3, 1219–53. What happens when we can add marbles with new colors to the Pólya urn as we go? The urn model maps onto a network-growth model.
- [223] N. Helfand, “Polya’s Urn and the Beta-Bernoulli process,” *University of Chicago REU 2013* (2013) n.p. <http://math.uchicago.edu/~may/REU2013/>.
- [224] C. M. Caves, C. A. Fuchs, and R. Schack, “Unknown quantum states: The quantum de Finetti representation,” *Journal of Mathematical Physics* **43** (2002) 4357, [arXiv:quant-ph/0104088](#).

- [225] J. Aczél, B. Forte, and C. T. Ng, “Why the Shannon and Hartley entropies are ‘natural’,” *Advances in Applied Probability* **6** (1974) no. 1, 131–46, [JSTOR:1426210](#).
- [226] J. C. Baez, T. Fritz, and T. Leinster, “A characterization of entropy in terms of information loss,” *Entropy* **13** (2011) no. 11, 1945–57, [arXiv:1106.1791 \[cs.IT\]](#).
- [227] G. Gbur, *Mathematical Methods for Optical Physics and Engineering*. Cambridge University Press, 2011.
<http://skullsinthestars.com/mathematical-methods-for-optics/>.
- [228] H. S. Wilf, *generatingfunctionology*. Academic Press, 1994.
- [229] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*. Springer, 1996.
- [230] B. Allen, M. A. Nowak, and U. Dieckmann, “Adaptive dynamics with interaction structure,” *The American Naturalist* **181** (2013) no. 6, E139–63. One of the papers on which I leaned greatly.
- [231] C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak, “Strategy selection in structured populations,” *Journal of Theoretical Biology* **259** (2009) no. 3, 570–81, [PMC:2710410](#).
- [232] N. Champagnat and A. Lambert, “Evolution of discrete populations and the canonical diffusion of adaptive dynamics,” *The Annals of Applied Probability* **17** (2007) no. 1, 102–55. <http://projecteuclid.org/euclid.aoap/1171377179>.
- [233] J. Van Cleve, “Social evolution and genetic interactions in the short and long term,” *bioRxiv* (2014) n.p.
- [234] C. E. Tarnita and P. D. Taylor, “Measures of relative fitness of social behaviors in finite structured population models,” *The American Naturalist* **184** (2014) no. 4, 477–88. <http://scholar.princeton.edu/sites/default/files/ctarnita/files/tarnitataylor2014.pdf>.
- [235] M. Doebeli, C. Hauert, and T. Killingback, “The evolutionary origin of cooperators and defectors,” *Science* **306** (2004) 859–62.
- [236] C. E. Tarnita, N. Wage, and M. A. Nowak, “Multiple strategies in structured populations,” *PNAS* **108** (2011) no. 6, 2334–37.
- [237] A. McAvoy and C. Hauert, “Structure coefficients and strategy selection in multiplayer games,” *Journal of Mathematical Biology* (2015) epub ahead of print.
- [238] I. M. Hastings, P. G. Bray, and S. A. Ward, “A requiem for chloroquine,” *Science* **298** (2002) no. 5591, 74–75.
- [239] R. L. Summers *et al.*, “Diverse mutational pathways converge on saturable chloroquine transport via the malaria parasite’s chloroquine resistance transporter,” *PNAS* **111** (2014) no. 17, E1759–67.
- [240] B. C. Kirkup and M. A. Riley, “Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors *in vivo*,” *Nature* **428** (2004) 412–14.
- [241] B. Wu, A. Traulsen, and C. S. Gokhale, “Dynamic properties of evolutionary multi-player games in finite populations,” *Games* **4** (2013) no. 2, 182–99.
- [242] A. Wagner, “Robustness and evolvability: a paradox resolved,” *Proceedings of the Royal Society B* **275** (2008) 91–100, [PMC:2562401](#).

Bibliography

- [243] R. Partha and K. Raman, “Revisiting robustness and evolvability: Evolution in weighted genotype spaces,” *PLoS One* **9** (2014) no. 11, e112792.
- [244] V. Janjić, R. Sharan, and N. Pržulj, “Modelling the yeast interactome,” *Scientific Reports* **4** (2014) 4273.
- [245] H. A. Orr, “The population genetics of beneficial mutations,” *Philosophical Transactions of the Royal Society B* **365** (2010) no. 1544, 1195–1201.
- [246] R. M. May, “Will a large complex system be stable?,” *Nature* **238** (1972) 413–14.
- [247] S. N. Majumdar and G. Schehr, “Top eigenvalue of a random matrix: large deviations and third order phase transition,” *Journal of Statistical Mechanics* **2014** (2014) P01012, [arXiv:1311.0580 \[cond-mat.stat-mech\]](#).
- [248] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Oxford University Press, 2006.
<http://pages.physics.cornell.edu/~sethna/StatMech/>.
- [249] A. B. Zamolodchikov, ““Irreversibility” of the flux of the renormalization group in a 2D field theory,” *JETP Letters* **43** (1986) no. 12, 730–32.
- [250] R. J. Creswick, H. A. Farach, and C. P. Poole, *Introduction to Renormalization Group Methods in Physics*. John Wiley & Sons, 1992.
- [251] W. D. McComb, *Renormalization Methods: A Guide for Beginners*. Oxford University Press, 2003.
- [252] A. Zee, *Quantum Field Theory in a Nutshell*. Princeton University Press, 2003.
- [253] Y. Bar-Yam, *Dynamics of Complex Systems*. Westview Press, 2003.
<http://necsi.edu/publications/dcs/index.html>.
- [254] J. Cardy, *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, 1996.
- [255] M. Doi, “Second quantization representation for classical many-particle system,” *Journal of Physics A* **9** (1976) no. 9, 1465–77.
- [256] M. Doi, “Stochastic theory of diffusion-controlled reaction,” *Journal of Physics A* **9** (1976) no. 9, 1479–95.
- [257] D. C. Mattis and M. L. Glasser, “The uses of quantum field theory in diffusion-limited reactions,” *Reviews of Modern Physics* **70** (1998) no. 3, 979–1001.
- [258] J. C. Baez and J. Biamonte, “A course on quantum techniques for stochastic processes,” [arXiv:1209.3632 \[quant-ph\]](#). An introductory text which starts in roughly the same place as our presentation but goes in a different direction after a while.
- [259] J. Baez and J. Dolan, “From finite sets to Feynman diagrams,” in *Mathematics Unlimited – 2001 and Beyond*, B. Engquist and W. Schmid, eds., pp. 29–50. Springer, 2001. [arXiv:math/0004133](#).
- [260] U. C. Täuber, M. Howard, and B. P. Vollmayr-Lee, “Applications of field-theoretic renormalization group methods to reaction-diffusion problems,” *Journal of Physics A* **38** (2005) no. 17, R79, [arXiv:cond-mat/0501678](#).
- [261] B. P. Vollmayr-Lee, “Field theory approach to diffusion-limited reactions,” Boulder school for condensed matter and materials physics, 2009.
<http://boulderschool.yale.edu/2009/boulder-school-2009-lecture-notes>.

- [262] M. Mobilia, I. T. Georgiev, and U. C. Täuber, “Spatial stochastic predator-prey models,” [arXiv:q-bio/0609039](#).
- [263] H. Abarbanel and J. B. Bronzan, “Structure of the Pomeranchuk singularity in Reggeon field theory,” *Physical Review D* **9** (1974) no. 8, 2397–410.
- [264] P. Grassberger and K. Sundermeyer, “Reggeon field theory and Markov processes,” *Physics Letters* **77B** (1978) no. 2, 220–22.
- [265] J. L. Cardy and R. L. Sugar, “Directed percolation and Reggeon field theory,” *Journal of Physics A* **13** (1980) L423–27.
- [266] A. Connes and D. Kreimer, “Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem,” *Communications in Mathematical Physics* **210** (2000) no. 1, 249–73, [arXiv:hep-th/9912092](#).
- [267] S. R. Broadbent and J. M. Hammersley, “Percolation processes I. Crystals and mazes,” *Mathematical Proceedings of the Cambridge Philosophical Society* **53** (1957) no. 3, 629–41.
- [268] K. A. Takeuchi, M. Kuroda, H. Chaté, and M. Sano, “Experimental realization of directed percolation criticality in turbulent liquid crystals,” *Physical Review E* **80** (2009) 051116, [arXiv:0907.4297 \[cond-mat\]](#).
- [269] H. Kelker and B. Scheurle, “A liquid-crystalline (nematic) phase with a particularly low solidification point,” *Angewandte Chemie* **8** (1969) no. 11, 884–85.
- [270] F. Peruani and C. F. Lee, “Fluctuations and the role of collision duration in reaction-diffusion systems,” *Europhysics Letters* **102** (2013) no. 5, 58001, [arXiv:1305.4466 \[cond-mat.soft\]](#).
- [271] J. C. Baez, B. C. Stacey, *et al.*, “Quantum techniques for reaction networks,” *Azimuth* (2013) n.p. <https://johncarlosbaez.wordpress.com/2013/06/11/quantum-techniques-for-reaction-networks/>.
- [272] H. K. Janssen, “On the nonequilibrium phase transition in reaction-diffusion systems with an absorbing stationary state,” *Zeitschrift für Physik B* **42** (1981) no. 2, 151–54.
- [273] P. Grassberger, “On phase transitions in Schlögl’s second model,” *Zeitschrift für Physik B* **47** (1982) no. 4, 365–74.
- [274] F. van Wijland, “Field theory for reaction-diffusion processes with hard-core particles,” *Physical Review E* **63** (2001) 022101, [arXiv:cond-mat/0010491](#).
- [275] N. Stollenwerk and V. Jansen, *Population Biology and Criticality: From critical birth-death processes to self-organized criticality in mutation pathogen systems*. Imperial College Press, 2011.
- [276] P. Grassberger and A. de la Torre, “Reggion field theory (Schlögl’s first model) on a lattice: Monte Carlo calculations of critical behaviour,” *Annals of Physics* **122** (1979) no. 2, 373–96.
- [277] A. G. Moreira and R. Dickman, “Critical dynamics of the contact process with quenched disorder,” *Physical Review E* **54** (1996) R3090, [arXiv:cond-mat/9604148](#).
- [278] H. K. Janssen, “Renormalized field theory of the Gribov process with quenched disorder,” *Physical Review E* **55** (1997) 6253.
- [279] J. Hooyberghs, F. Iglói, and C. Vanderzande, “Absorbing state phase transitions with quenched disorder,” *Physical Review E* **69** (2004) 066140, [arXiv:cond-mat/0402086](#).

Bibliography

- [280] K. Tainaka, “Vortices and strings in a model ecosystem,” *Physical Review E* **50** (1994) no. 5, 3401–09.
- [281] A. Dobrinevski, M. Alava, T. Reichenbach, and E. Frey, “Mobility-dependent selection of competing strategy associations,” *Physical Review E* **89** (2014) 012721, [arXiv:1401.1755 \[q-bio.PE\]](https://arxiv.org/abs/1401.1755).
- [282] J. Fox, “Zombie ideas in ecology: r and k selection,” *Oikos blog* (2011) n.p. <http://oikosjournal.wordpress.com/2011/06/29/zombie-ideas-in-ecology-r-and-k-selection/>.
- [283] B. C. Stacey, “Adaptive networks,” *Science After Sunclipse* (2011) n.p. <http://www.sunclipse.org/?p=920>.
- [284] M. Pascual, M. Roy, and K. Laneri, “Simple models for complex systems: exploiting the relationship between local and global densities,” *Theoretical Ecology* **4** (2011) no. 2, 211–22.
- [285] C. R. Darwin, *The Origin of Species*. John Murray, sixth ed., 1872. http://darwin-online.org.uk/EditorialIntroductions/Freeman_OntheOriginofSpecies.html.
- [286] C. R. Darwin, *The Descent of Man, and Selection in Relation to Sex*. John Murray, 1871. http://darwin-online.org.uk/EditorialIntroductions/Freeman_TheDescentofMan.html.
- [287] A. Stoltzfus and K. Cable, “Mendelian–Mutationism: The Forgotten Evolutionary Synthesis,” *Journal of the History of Biology* **47** (2014) no. 4, 501–46.
- [288] P. Z. Myers, “The problem with evo-devo,” *Pharyngula* (2012) n.p. <http://freethoughtblogs.com/pharyngula/2012/02/21/the-problem-with-evo-devo/>.
- [289] J. R. Madden *et al.*, “Local traditions of bower decoration by spotted bowerbirds in a single population,” *Animal Behaviour* **68** (2004) no. 4, 759–65.
- [290] M. van Veelen, S. Luo, and B. Simon, “A simple model of group selection that cannot be analyzed with inclusive fitness,” *Journal of Theoretical Biology* **360** (2014) 279–89. http://math.ucdenver.edu/~bsimon/vanveelen_luo_simon_jtb_2014.pdf.
- [291] J. N. Pruitt and C. J. Goodnight, “Site-specific group selection drives locally adapted group compositions,” *Nature* **514** (2014) 359–62.
- [292] C. Goodnight, “Multilevel selection theory and evidence: a critique of Gardner, 2015,” *Journal of Evolutionary Biology* **28** (2015) epub ahead of print.
- [293] C. J. Goodnight, “Gardner’s theory of multilevel selection: Where he goes wrong and why,” *Evolution in Structured Populations* (2015) n.p. <http://blog.uvm.edu/cgoodnig/2015/01/28/gardners-theory-of-multilevel-selection-where-he-goes-wrong-and-why/>.
- [294] C. J. Goodnight, “A one line proof of Fisher’s fundamental theorem,” *Evolution in Structured Populations* (2015) n.p. <http://blog.uvm.edu/cgoodnig/2015/01/21/a-one-line-proof-of-fishers-fundamental-theorem/>. Presents the result as a “truism” (or as I called it, an “arithmetic identity”).
- [295] M. van Veelen, “On the use of the Price equation,” *Journal of Theoretical Biology* **237** (2005) no. 4, 412–26, PMID:15953618.

- [296] M. van Veelen, J. Garcia, M. W. Sabelis, and M. Egas, “Group selection and inclusive fitness are *not* equivalent; the Price equation vs. models and statistics,” *Journal of Theoretical Biology* **299** (2012) 64–80, PMID:21839750.
- [297] W. D. Hamilton, “The genetical evolution of social behaviour. I,” *Journal of Theoretical Biology* **7** (1964) no. 1, 1–16.
- [298] C. Johnston, “Biological warfare flares up again between EO Wilson and Richard Dawkins,” *The Guardian* (2014) n.p. <http://www.theguardian.com/science/2014/nov/07/richard-dawkins-labelled-journalist-by-eo-wilson>.
- [299] E. O. Wilson and B. Hölldobler, “Eusociality: Origin and consequences,” *Proceedings of the National Academy of Sciences* **102** (2005) no. 38, 13367–71, PMC:1224642.
- [300] B. Hölldobler and E. O. Wilson, *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*. W. W. Norton, 2009.
- [301] B. Hölldobler, “Multi-component signals in ant communication,” in *Social Behaviour: Genes, Ecology and Evolution*, T. Székely, A. J. Moore, and J. Komdeur, eds., pp. 127–31. Cambridge University Press, 2010.
- [302] V. C. Wynne-Edwards, *Animal Dispersion in Relation to Social Behavior*. Hafner, 1962.
- [303] G. B. Pollock, “Suspending disbelief—of Wynne-Edwards and his reception,” *Journal of Evolutionary Biology* **2** (1989) 205–21. D. S. Wilson pointed me to this article.
- [304] V. C. Wynne-Edwards, “Intergroup selection in the evolution of social systems,” *Nature* **200** (1963) 623–26.
- [305] S. Wright, “Evolution in a Mendelian population,” *The Anatomical Record* **44** (1929) no. 3, 287. Abstract of a presentation.
- [306] S. Wright, “Evolution in Mendelian populations,” *Genetics* **16** (1931) no. 2, 97–159, PMC:1201091.
- [307] J. B. S. Haldane, *The Causes of Evolution*. Longmans, Green & Co. Limited, 1932.
- [308] A. H. Sturtevant, “Essays on evolution. II. On the effects of selection on social insects,” *The Quarterly Review of Biology* **13** (1938) no. 1, 74–76, JSTOR:2808460.
- [309] G. C. Williams and D. C. Williams, “Natural selection of individually harmful social adaptations among sibs with special reference to social insects,” *Evolution* **11** (1957) no. 1, 32–39, JSTOR:2405809.
- [310] J. Maynard Smith, “Group selection and kin selection,” *Nature* **201** (1964) 1145–47. A critique of Wynne-Edwards that attributed to him an image of population structure he probably did not have in mind.
- [311] B. Simon, “Continuous-time models of group selection, and the dynamical insufficiency of kin selection models,” *Journal of Theoretical Biology* **349** (2014) 22–31. <http://www.math.ucdenver.edu/~bsimon/ContTimeModelsGroupSelection.pdf>.
- [312] E. M. Johnson, “The good fight,” *Scientific American Blogs: The Primate Diaries* (2012) n.p. <http://blogs.scientificamerican.com/primate-diaries/2012/07/09/the-good-fight/>.
- [313] L. Gonick, *The Cartoon Guide to (Non)Communication*. HarperPerennial, 1993. Reflections on how to communicate clearly about complicated topics, by a professional.

Bibliography

- [314] J. C. Baez, “Network theory (part 4),” *Azimuth* (2011) n.p.
<http://johncarlosbaez.wordpress.com/2011/04/06/network-theory-part-4/>.
- [315] E. Ziv, M. Middendorf, and C. Wiggins, “An information-theoretic approach to network modularity,” *Physical Review E* **71** (2005) 046117, [arXiv:q-bio.QM/0411033](#).
- [316] S. Fortunato, “Community detection in graphs,” *Physics Reports* **486** (2010) 75–174, [arXiv:0906.0612 \[physics.soc-ph\]](#).
- [317] B. H. Good, Y.-A. de Montjoie, and A. Clauset, “The performance of modularity maximization in practical contexts,” *Physical Review E* **81** (2009) 046106, [arXiv:0910.0165 \[physics.data-an\]](#).
- [318] R. K. Merton, “The Matthew Effect in science,” *Science* **159** (1968) 56–63.
- [319] D. de Solla Price, “A general theory of bibliometric and other cumulative advantage processes,” *Journal of the American Society for Information Science* **27** (1976) 292–306.
- [320] A. L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science* **286** (1999) 509–12, [arXiv:cond-mat/9910332](#).
- [321] R. Albert, H. Jeong, and A. L. Barabási, “Attack and error tolerance of complex networks,” *Nature* **406** (2000) 378–82.
- [322] N. Slonim and L. Tishby, “Agglomerative information bottleneck,” *Advances in Neural Information Processing Systems (NIPS)* **12** (1999) 617–23.
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.122.8863>.
- [323] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
- [324] N. D. Mermin, “Hidden variables and the two theorems of John Bell,” *Reviews of Modern Physics* **65** (1993) no. 3, 803–15.
- [325] R. W. Spekkens, “Evidence for the epistemic view of quantum states: A toy theory,” *Physical Review A* **75** (2007) 032110, [arXiv:quant-ph/0401052](#).
- [326] S. J. van Enk, “A toy model for quantum mechanics,” *Foundations of Physics* **37** (2007) 1447–60, [arXiv:0705.2742 \[quant-ph\]](#).
- [327] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, “Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction,” *Physical Review A* **86** (2012) 012103, [arXiv:1111.5057 \[quant-ph\]](#).
- [328] R. W. Spekkens, “Quasi-quantization: classical statistical theories with an epistemic restriction,” [arXiv:1409.5041 \[quant-ph\]](#).
- [329] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, “The resource theory of stabilizer computation,” *New Journal of Physics* **16** (2014) 013009, [arXiv:1307.7171 \[quant-ph\]](#).
- [330] J. C. Baez, E. Lerman, *et al.*, “Network theory (part 17),” *Azimuth* (2011) n.p.
<http://johncarlosbaez.wordpress.com/2011/11/12/network-theory-part-17/>.
- [331] J. Morton, “Categorified algebra and quantum mechanics,” *Theory and Applications of Categories* **16** (2006) no. 29, 785, [arXiv:math/0601458 \[math.QA\]](#).
- [332] D. E. Knuth, *Surreal Numbers*. Addison–Wesley, 1974.
- [333] J. H. Conway and R. K. Guy, *The Book of Numbers*. Springer–Verlag, 1996.

- [334] A. Joyal, “Remarques sur la théorie des jeux à deux personnes,” *Gazette des sciences mathématiques du Québec* **1** (1977) no. 4, 46–52.
- [335] F. M. Cross, “The evolution of the Proto-Canaanite alphabet,” *Bulletin of the American Schools of Oriental Research* (1954) no. 134, 15–24, [JSTOR:1355624](#).
- [336] Anon., “U, n.1,” in *Oxford English Dictionary*. Oxford University Press, 1989.
- [337] Q. D. Atkinson and R. D. Gray, “Curious parallels and curious connections—Phylogenetic thinking in biology and historical linguistics,” *Systematic Biology* **54** (2005) no. 4, 513–26.
- [338] B. Allen, M. Kon, and Y. Bar-Yam, “A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats,” *The American Naturalist* **174** (2009) 236–43. <http://necsi.edu/headlines/amnat2.html>.
- [339] T. Leinster and C. A. Cobbold, “Measuring diversity: the importance of species similarity,” *Ecology* **93** (2012) no. 3, 477–89. <http://johncarlosbaez.wordpress.com/2011/11/07/measuring-biodiversity/>.
- [340] A. Chao, C.-H. Chiu, and L. Jost, “Phylogenetic diversity measures based on Hill numbers,” *Philosophical Transactions of the Royal Society B* **365** (2010) 3599–609, [PMC:2982003](#).
- [341] L. Jost, “Entropy and diversity,” *Oikos* **113** (2006) no. 2, 363–75. <http://www.loujost.com/Statistics%20and%20Physics/Diversity%20and%20Similarity/JostEntropy%20AndDiversity.pdf>.
- [342] H. Sayama and J. Akaishi, “Characterizing interdisciplinarity of researchers and research topics using web search engines,” *PLOS ONE* **7** (2012) no. 6, e38747, [arXiv:1201.3592 \[cs.SI\]](#).
- [343] C. Ricotta and L. Szeidl, “Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao’s quadratic index,” *Theoretical Population Biology* **70** (2006) 237–43.
- [344] E. Prugovečki, “Information-theoretical aspects of quantum measurement,” *International Journal of Theoretical Physics* **16** (1977) no. 5, 321–31.
- [345] G. M. D’Ariano, P. Perinotti, and M. F. Sacchi, “Informationally complete measurements and group representation,” *Journal of Optics B* **6** (2004) no. 6, S487, [arXiv:quant-ph/0310013](#).
- [346] C. A. Fuchs, “Quantum mechanics as quantum information (and only a little more),” in *Quantum Theory: Reconsideration of Foundations*, A. Khrennikov, ed., pp. 463–543. Växjö University Press, 2002. [arXiv:quant-ph/0205039](#).
- [347] G. Zauner, *Quantum Designs – Foundations of a Noncommutative Theory of Designs*. PhD thesis, University of Vienna, 1999. <http://www.gerhardzauner.at/qdmye.html>. First published appearance of SICs in quantum physics.
- [348] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, “Symmetric informationally complete quantum measurements,” *Journal of Mathematical Physics* **45** (2004) no. 6, 2171, [arXiv:quant-ph/0310075](#).
- [349] C. A. Fuchs and R. Schack, “Quantum-Bayesian coherence,” *Reviews of Modern Physics* **85** (2013) no. 4, 1693–1715, [arXiv:1301.3274 \[quant-ph\]](#).
- [350] A. J. Scott and M. Grassl, “SIC-POVMs: A new computer study,” *Journal of Mathematical Physics* **51** (2010) 042203, [arXiv:0910.5784 \[quant-ph\]](#).

Bibliography

- [351] D. M. Appleby, H. B. Dang, and C. A. Fuchs, “Symmetric informationally-complete quantum states as analogues to orthonormal bases and minimum-uncertainty states,” *Entropy* **16** (2014) 1484–92, [arXiv:0707.2071 \[quant-ph\]](#).
- [352] S. T. Flammia. Unpublished, 2004.
- [353] N. S. Jones and N. Linden, “Parts of quantum states,” *Physical Review A* **71** (2005) no. 1, 012324, [arXiv:quant-ph/0407117](#).
- [354] D. M. Appleby, S. T. Flammia, and C. A. Fuchs, “The Lie algebraic significance of symmetric informationally complete measurements,” *Journal of Mathematical Physics* **52** (2011) 022202, [arXiv:1001.0004 \[quant-ph\]](#).
- [355] C. A. Fuchs and R. Schack, “A Quantum-Bayesian route to quantum-state space,” *Foundations of Physics* **41** (2011) no. 3, 345–56, [arXiv:0912.4252 \[quant-ph\]](#).
- [356] D. M. Appleby, Å. Ericsson, and C. A. Fuchs, “Properties of QBist state spaces,” *Foundations of Physics* **41** (2011) no. 3, 564–79, [arXiv:0910.2750 \[quant-ph\]](#).
- [357] C. A. Fuchs and B. C. Stacey, “Some negative remarks on operational approaches to quantum theory,” in *Quantum Theory: Informational Foundations and Foils*, G. Chiribella and R. W. Spekkens, eds. Springer, 2015. [arXiv:1401.7254 \[quant-ph\]](#). Edited transcript of a 2011 lecture. Among other things, I fixed the introductory joke.
- [358] G. N. Tabia, “Experimental scheme for qubit and qutrit SIC-POVMs using multiport devices,” *Physical Review A* **86** (2012) no. 6, 062107, [arXiv:1207.6035 \[quant-ph\]](#).
- [359] G. N. M. Tabia and D. M. Appleby, “Exploring the geometry of qutrit state space using symmetric informationally complete probabilities,” *Physical Review A* **88** (2013) no. 1, 012131, [arXiv:1304.8075 \[quant-ph\]](#).
- [360] I. Bengtsson, K. Blanchfield, and A. Cabello, “A Kochen–Specker inequality from a SIC,” *Physics Letters A* **376** (2010) no. 4, 374–76, [arXiv:1109.6514 \[quant-ph\]](#).
- [361] H. B. Dang, K. Blanchfield, I. Bengtsson, and D. M. Appleby, “Linear dependencies in Weyl–Heisenberg orbits,” *Quantum Information Processing* **12** (2012) no. 11, 3449–75, [arXiv:1211.0215 \[quant-ph\]](#).
- [362] D. M. Appleby, I. Bengtsson, and S. Chaturvedi, “Spectra of phase point operators in odd prime dimensions and the extended Clifford group,” *Journal of Mathematical Physics* **49** (2008) 012102, [arXiv:0710.3013 \[quant-ph\]](#).
- [363] B. C. Stacey, “SIC-POVMs and compatibility among quantum states,” [arXiv:1404.3774 \[quant-ph\]](#).
- [364] J. Fox, *The Myth of the Rational Market*. HarperCollins, 2009.
- [365] P. Krugman, “Irregular economics,” *The Conscience of a Liberal* (2011) n.p. <http://krugman.blogs.nytimes.com/2011/08/25/irregular-economics/>. “What,” Krugman asks, “is the compelling evidence that the vision of a competitive, efficient economy allocating resources to the right uses is actually a good description of the world we live in?” The attempt to explain business cycles with “regular economics” has been “a dismal failure, even if the practitioners refuse to admit it,” he notes.
- [366] M. Lagi, Y. Bar-Yam, K. Z. Bertrand, and Y. Bar-Yam, “The food crises: A quantitative model of food prices including speculators and ethanol conversion,” Tech. Rep. 2011-09-01, NECSI, 2011. [arXiv:1109.4859 \[q-fin.GN\]](#). <http://necsi.edu/research/social/foodprices.html>.

- [367] M. Lagi and Y. Bar-Yam, “The European debt crisis: Defaults and market equilibrium,” Tech. Rep. 2012-09-27, NECSI, 2012. [arXiv:1209.6369 \[q-fin.GN\]](#). <http://necsi.edu/research/economics/bondprices/>.
- [368] M. Lagi, K. Z. Bertrand, and Y. Bar-Yam, “Food security and political instability: From ethanol and speculation to riots and revolutions,” in *Conflict and Complexity*. Springer, 2015.
- [369] P. N. Malaney, *The Index Number Problem: A Differential Geometric Approach*. PhD thesis, Harvard University, 1996.
- [370] L. Smolin, “Time and symmetry in models of economic markets,” [arXiv:0902.4274 \[q-fin.GN\]](#).
- [371] J. C. Baez *et al.*, “Mathematical economics,” *Azimuth* (2011) n.p. <https://johncarlosbaez.wordpress.com/2011/01/08/mathematical-economics/>.
- [372] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, “Duplication-divergence model of protein interaction network,” *Physical Review E* **71** (2005) 061911, [PMC:2092385](#).
- [373] H. Avalos, *The End of Biblical Studies*. Prometheus, 2007.
- [374] R. P. Feynman, *QED: The Strange Theory of Light and Matter*. Princeton University Press, 1985. <http://press.princeton.edu/chapters/i2352.html>.
- [375] A. Pais, “Einstein and the quantum theory,” *Reviews of Modern Physics* **51** (1979) 863–914.
- [376] B. Switek, “John Scopes and textbook cardboard,” *Laelaps* (2008) n.p. <http://scienceblogs.com/laelaps/2008/05/21/john-scopes-and-textbook-cardb/>. Not the first author to use the term “textbook cardboard,” but the one who introduced me to it.
- [377] W. Thomson, “Nineteenth century clouds over the dynamical theory of heat and light,” *Philosophical Magazine, Sixth Series* **2** (1901) 1–40. Lord Kelvin’s thoughts on outstanding conceptual troubles at the turn of the twentieth.
- [378] P. G. Tait, “On the foundations of the kinetic theory of gases,” *Transactions of the Royal Society of Edinburgh* **33** (1886) 14 May.
- [379] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*. Charles Scribner’s Sons, 1901. <http://www.aip.org/history/gap/Gibbs/Gibbs.html>.
- [380] J. C. Maxwell, “Illustrations of the dynamical theory of gases,” *Philosophical Magazine, Fourth Series* **20** (1860) 21–37.
- [381] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*. Addison–Wesley, 1964. http://www.feynmanlectures.caltech.edu/I_40.html#Ch40-S6.
- [382] B. C. Stacey, “More decimal digits,” *Science After Sunclipse* (2011) n.p. <https://www.sunclipse.org/?p=947>. The story of how I found a mistake in the *Feynman Lectures*.
- [383] D. Howard, “Who invented the ‘Copenhagen Interpretation’? A study in mythology,” *Philosophy of Science* **71** (2004) no. 5, 669–82.
- [384] K. Camilleri, “Constructing the myth of the Copenhagen Interpretation,” *Perspectives in Science* **17** (2009) no. 1, 26–57. <http://muse.jhu.edu/journals/posc/summary/v017/17.1.camilleri.html>.

Bibliography

- [385] B. C. Stacey, “Von Neumann was not a Quantum Bayesian,” [arXiv:1412.2409 \[physics.hist-ph\]](#). Twenty-four pages, written in response to one sentence on Wikipedia.
- [386] K. Camilleri and M. Schlosshauer, “Niels Bohr as philosopher of experiment: Does decoherence theory challenge Bohr’s doctrine of classical concepts?,” *Studies in History and Philosophy of Modern Physics* **49** (2015) 73–83, [arXiv:1502.06547 \[quant-ph\]](#). The title is an instance of Betteridge’s Law: Any headline that ends in a question mark can be answered with a *No*.
- [387] J. Cham, “Writing your thesis outline,” *PhD Comics* (2006) n.p. <http://www.phdcomics.com/comics/archive.php?comicid=715>.
- [388] M. Kardar and L. Mirny, “Statistical physics in biology, Spring 2005,” *MIT OpenCourseWare* (2005) n.p. <http://dspace.mit.edu/handle/1721.1/81298>. Course notes for the class where I learned about the Fokker–Planck equation and its application to population genetics.
- [389] D. Hoffman, “The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age [book review],” *Notices of the AMS* **45** (1998) no. 2, 260–67. <http://www.ams.org/notices/199802/bookrev-hoffman.pdf>. Worth it for the remarks about “video proofs” alone.
- [390] F. R. Blattner *et al.*, “The complete genome sequence of *Escherichia coli* K-12,” *Science* **277** (1997) no. 5331, 1453–62.
- [391] A. G. Riess *et al.*, “Observational evidence from supernovae for an accelerating universe and a cosmological constant,” *The Astronomical Journal* **116** (1998) no. 3, 1009–38, [arXiv:astro-ph/9805201](#).
- [392] S. Perlmutter *et al.*, “Measurements of Ω and Λ from 42 high-redshift supernovae,” *The Astrophysical Journal* **517** (1999) no. 2, 565–86, [arXiv:astro-ph/9812133](#).