7

Life II: Developmental Biology—
Complex by Design

Conceptual Outline

7.1 Developmental biology strives to understand the sequence of events by which a single cell becomes a system of many differentiated interacting cells. This process involves placing different structures in particular locations and interconnecting them.

7.2 To model differentiation we focus on the formation of color patterns on animal skins that have a variety of forms. Cellular automaton models show the relevance of local activation and long-range inhibition of pigment production to the formation of patterns. Chemical reaction-diffusion systems illustrate similar patterns using slow- and fast-diffusing species.

7.3 Other elements of the tool kit for developmental processes include mechanisms for changes in cell structure, cell motion, timing and counting. Of particular interest are sequential steps (programs) that can form branching structures.

7.4 Theoretical modeling can better complement phenomenological studies of biological systems if the different objectives of theory and experiment are recognized.

7.5 The approach of developmental biology to the design of complex systems may be a useful framework for considering the design of complex artificial systems.

7.6 Models of pattern formation may be better suited to discussions of global properties of the evolution of organisms than the models discussed in Chapter 6.

7.1 Developmental Biology: Programming a Brick

Reproduction in multicellular organisms, animals and plants, occurs through a process of development from a single cell. The fundamental objective of developmental biology is to understand how an individual cell through cell division, differentiation and growth results in a complex physiology. The controls for this process of

development are present within the initial cell and also in the environment in which the cell develops.

Our concern in this chapter is largely with the cellular behavior in development rather than with the internal functioning of the cell. However, in the following paragraphs we discuss briefly models for the mechanisms that exercise control over the developmental process as part of the internal functioning of the cell.

It is generally believed that the design of plant or animal physiology is contained within the nuclear DNA of the cell.DNA is often called the blueprint for the biological organism. However, it is clear that DNA does not function like an architect's blueprint because the information does not represent the structure of the physiology in a direct way—there is no homunculus there. For our purposes it is convenient to think about the DNA blueprint as a program that specifies the interaction between a cell and its environment,including cells in its vicinity, as well as the internal functioning of the cell. However, in describing DNA as a program we are implicitly subsuming the functions and description of the entire cellular machinery in the DNA. For our abstract purposes, there is no difference in various sources of information,as there is no essential difference between information that is found on the tape of a Turing machine and information in the table of the read-write head (see Section 1.9.4). There are, however, other conceptual issues to address.

First, we must clarify the nature of DNA function within the cell.DNA serves at least in part as a collection of templates (genes) that may be thought of as blueprints for protein chains. These templates are sometimes being transcribed (active) and sometimes not being transcribed (inactive). Thus, the role of DNA at a particular time is described by a set of transcription activities. The activity of a particular gene depends on the activity of other genes. Thus, a useful analogy may be a neural network model where the transcription activities are analogous to the neuronal activities in the network. Like the synapses of the network, the molecular machinery of the cell mediates the activities (and performs the transcriptions) of the DNA. The patterns of activity of the transcription of DNA are a part of the patterns of activity of the cell as a whole which constitute possible behaviors of the cell. Thus it may be reasonable to consider the relevance of attractors of patterns of activity, as in the neural network models discussed in Chapter 2, to the study of cellular function. The development of an organism consists of a temporal sequence of such patterns of cellular function.

Second, we must clarify the relationship of information and behavior. It is likely that the DNA in a cell contains a large proportion of the information needed to describe the function of the cell, the developmental dynamics and the physiological function of the organism. However, this does not mean that the DNA should be thought of as controlling the processes in the conventional sense of the term "control." A useful analogy is the role of a library in society. It is quite likely that most of the information about the function of society in one way or another may be found in the Library of Congress. However, this does not mean that the library controls this function. It may, indeed, be better to think about the molecules in the cell as akin to a society of entities that act upon each other and respond to external stimuli.DNA then

serves this society as a source of information—in part as a repository of blueprints for the manufacture of cellular machinery.

In this regard it may be helpful, though somewhat subtle, to recognize that DNA is not by itself a complex organism. It does not satisfy our criteria of nondivisibility, since its structure and behavior (including transcription) is essentially local. It is only when the information in DNA takes form in the context of cellular or organismal behavior that the behavior is itself complex, and the system as a whole satisfies the conditions of a complex organism. Incidentally, this is also a reason that the structure of DNA does not satisfy the 7±2 rule—there are 23 pairs of homologue chromosomes in most human beings, and a wide variation in the number of chromosomes in other organisms.

Returning to our central focus in this chapter, for our purposes development is a largely deterministic sequence of cellular states that results in a multicellular organism. In this sense the organism can be described as the result of a program, since all deterministic processes can be so described. The program is largely contained in the original cell. It is essential to recognize that all cells of an organism begin from one cell in a unique state, and therefore inherit all or parts of the same set of information, and thus the same program.

The central problem of developmental biology is to describe how the cells differentiate in such a way as to place particular functions in particular locations in the body—not to describe the specific eventual function of each cell. Part of this problem is to describe how cells become interconnected by necessary structures formed out of individual cells such as long branching neurons, or many-celled structures such as blood vessels. This must be achieved by the program that specifies the sequence of cell states and cell interactions. The overall process of development is shown in Fig. 7.1.1.

Biological development is a systematic approach to the very difficult problem of designing complex systems. It enables the creation of a large variety of systems. In studying this approach it may be helpful to think about designing a building in a similar manner. Allowing some imagination, we might consider writing a program for a brick. The program describes how a brick should move and interact with other bricks in its vicinity. Providing the same program for each brick in a pile, we walk away and return to find the whole building, with windows, ducts, and utilities in place. Cells, unlike bricks, are themselves like organisms in consuming resources and producing waste; they are self-reproducing and mobile. They also have the ability to change shape. Through shape change and changes in chemical processes they can adopt a large variety of functions in a multicellular organism. Even if we endow bricks with similar abilities, it still requires careful thought to understand how the design of a complex structure can arise from a program describing their interactions.

It is significant that this approach balances design with self-organization. In Chapter 6 on evolution, we assumed a self-organizing process that occurred by chance and external selection. In contrast, organism development should reliably achieve a desired outcome from a preexisting (internal) design. Nevertheless, the built-in design directs a dynamic process where mutually interacting entities self-organize into the desired complex structure.

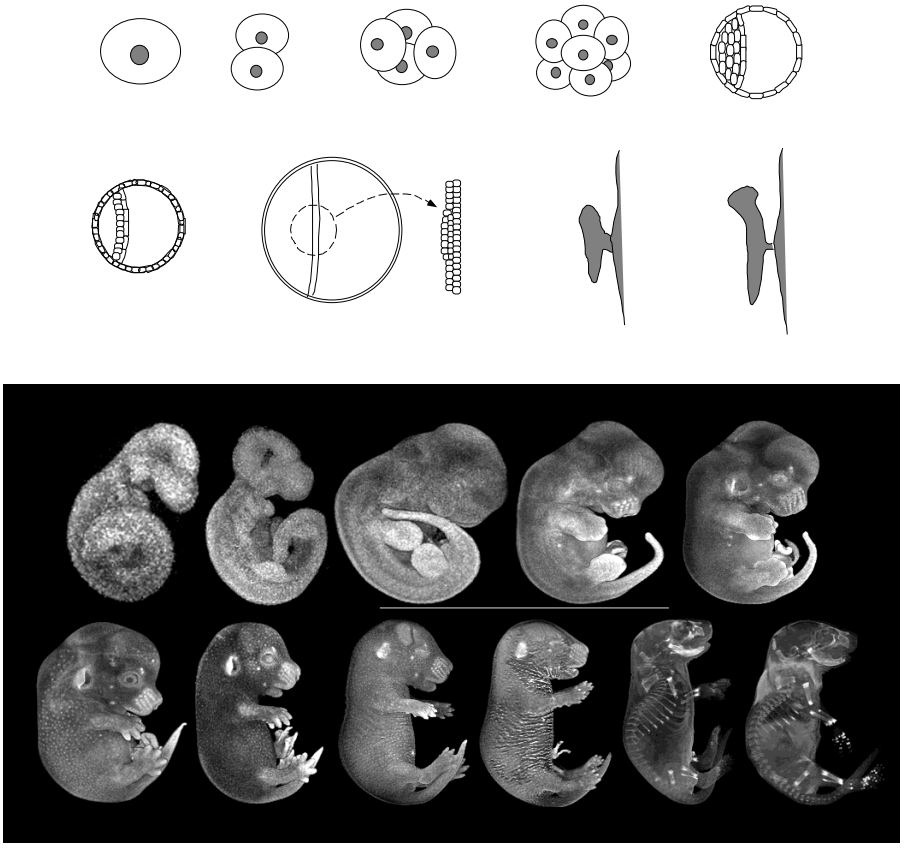


Figure 7.1.1 Illustration of some of the stages in the development of an animal. The top two rows are schematic illustrations of the initial stages where a single fertilized cell undergoes multiple divisions to form a spherical shell with a membrane separating its internal cavity into two parts that become the primary yolk sack and the amniotic cavity. Cells from part of the internal membrane then form the growing fetus. The bottom two rows are magnetic resonance microscopy images of mouse fetal development from 9 days to birth. All images are shown at the same size despite a 10-fold increase in fetal dimensions from the first to last image of this sequence. The multicellular structure of the organism arises through a set of programmed steps originating in a single cell. The identification of processes and mechanisms for this development is the subject of developmental biology (magnetic resonance microscopy images are courtesy of Brad Smith, Elwood Linney and the Center for In Vivo Microscopy at Duke University [A National Center for Research Resources, NIH]). ∎

For many who have had occasion to contemplate a newborn, development is miraculous. From a scientific point of view there are at least two reasons that this reaction arises. First, the relationship between process and outcome is emergent—the relationship between individual parts of the dynamics and the whole is difficult to understand. This is the nature of a complex self-organizing process. Second, designing a dynamic process that can reliably arrive at a specific complex outcome is difficult. When a process involves many steps and an error in any step may give rise to failure, the likelihood that the process will be successful is vanishingly small. Our analogy with a computer program is telling, since a single bit error in computer hardware or software would generally cause failure.It is useful to compare this with our discussion of protein folding in Chapters 4 and 5, where we were also concerned about arriving at a definite final structure. In Chapter 4 we considered exploration of conformation space to find an energy minimum. As long as the dynamics could reach the energy minimum, its identity was not in question. In Chapter 5 we argued that directed sequential steps could arrive at a desired final structure. Here we recognize that in a strictly directed (deterministic) process, there must be no error in the dynamics so that there will be no error in the eventual structure. What is particularly remarkable is that the dynamic process must at the same time be stable to many perturbations, and yet modifiable through mutations that enable evolutionary changes. To understand how this is possible we must eventually recognize that the dynamics as a whole must be formed out of a sequence of attractors that are sufficiently stable to be the outcome of a variety of intermediaries. In this way the nonequilibrium dynamics and its outcome may be relatively stable to perturbations.

From the most basic complex-systems point of view, the problem of developmental biology is composed out of two parts: first, to identify general and specific processes that cause a homogeneous set of cells to differentiate in a controlled fashion so that specific structures are located in specific locations with respect to each other; and second, to identify mechanisms for creating structures that interconnect or support various functional regions of the system. Much of the quantitative modeling of such processes is relatively recent. In this chapter we focus on the problem of differentiation. In Section 7.2 we describe models of the formation of patterns on animal skins. This problem captures an essential aspect of differentiation and structure. The advantage of such patterns is that their structure is not very specific and therefore lends itself to a simpler analysis. However, the interplay of such patterns with specific boundary conditions can give rise to well-defined structures when they are necessary in development. In Section 7.3 we describe some more tools necessary for developmental formation of physiological systems. Of particular emphasis is the formation of branching structures found in plants and animals in the lungs, nervous and vascular systems. In Section 7.4 we discuss some of the general objectives and methodologies of theory and mathematical modeling of biological systems. In Section 7.5 we discuss the general properties of organization by design in biological complex systems and contrast it with the conventional approaches used in human design and engineering. Finally, in Section 7.6 we return to consider the implications of the models of pattern formation in this chapter for the problem of evolution discussed in Chapter 6.

7.2 Differentiation: Patterns in Animal Colors

7.2.1 *Introduction to pigment patterns*

Many animals have patterns of coloration on their external surfaces. Color is the result of pigment produced in cells. Often the patterns are composed of only two different colors, but in some cases there are more. For our purposes,the examples that are convenient to think about are the patterns on the fur of both predator and prey mammals. Zebras, giraffes, tigers,leopards and many others have distinctive patterns as a species. These patterns also vary in more subtle ways from individual to individual. Other kinds of patterns are present in some insects—particularly butterflies—fish—particularly tropical fish—and birds—particularly tropical birds.

The functional relevance of patterns or brilliant coloration for animals is an interesting topic of study. We can try to understand the reasons for coloration through the concepts of evolution discussed in the previous chapter. Evolutionary theory suggests that such physiological attributes arise from a survival advantage. It is a common practice to offer explanations for the existence of physiologic or behavioral features based on this premise. The ultimate difficulty is that these explanations, no matter how well reasoned, are rarely subject to direct experimental test. However, there appears little doubt that a uniform color for some animals is used for camouflage within a well-defined environment. This is characteristic of various green, brown or black insects and lizards that are found on leaves, various tree trunks, or the ground. Patterns of coloration, whether of black and white or of brilliant colors,appear to be directly counter to this purpose. Alternative explanations rely upon some form of social or collective behavior. The coloration of prey such as zebras and giraffes might serve to confuse predators because, in the context of a herd of animals, it inhibits the distinction of one individual from another. The boundaries between animals become less distinct than the internal coloration boundaries. Since the herd as a whole is not readily attacked, the individual disguised as part of a larger system is protected. This is consistent with the general discussion in Chapter 6 about the nature of collective behavior. However, this does not explain the coloration in the predators—tigers,leopards, cheetahs, etc. The careful distinctness of the patterns of different species,however, suggests that they serve as identification. The ability to identify animals of the same species either for herding (animals finding their way back to the herd or smaller group) or for mating may be more important for survival than camouflage. It may also be that individuals—mates, young or others—are identified through the specific distinctions between individual coloring patterns. Regardless, the functional purpose of colors is not directly relevant to the problem of determining a process that can give rise to them—the topic of this chapter.

Why are color patterns interesting as a problem in developmental biology? It would seem that they are quite incidental to more important problems such as the formation of limbs, the development of organs and the formation of networks,neural or vascular. While coloration appears to be superficial, it captures a basic feature necessary for many of the other processes—differentiation. A central problem in development is to assign distinct tasks. In order for limbs to develop, at some point in time there must

be an identification of which cells are to proliferate in such a way as to give rise to the limbs, and which cells are not to proliferate. This requires the formation of a pattern in the initially undifferentiated cells. Only after a pattern has been established can the processes associated with differential function of the cells proceed. In a more general context, understanding pattern formation as a form of spatial and temporal structure is a central issue in the formation and function of complex systems in general.

Our objective is to construct mathematical models that can result in the formation of patterns such as those present on the skins of mammals (Fig. 7.2.1). The essentially two-dimensional animal surfaces enable us to illustrate more readily the models than if they were in three dimensions. The models might use a cellular space with a variable representing the color of each cell in an array. Since many of these animals have essentially two colors, we can use a binary variable s_i. This type of model is suggestive of a simple cellular automaton (CA, Section 1.5) where the individual cell determines its state (the color at that location) as a consequence of interactions with neighboring cells. Indeed, the process of intercellular influence in biology is generally suggestive of a CA—as long as communication between cells is local, and we do not consider migration of cells or changes in their shape. The most direct model represents each biological cell by a lattice cell; however, we can also consider a homogenous region of biological cells to be represented by a single lattice cell. Such CA are often natural models for processes that take us from the behavior of an individual cell (or homogenous region) to the inhomogenous behavior of a collection of cells. On a finer scale we can model the diffusion and reaction of chemical messengers between cells and their effect on pigmentation. This provides an additional level of detail to models of such patterns. In Section 7.2.2 we will consider CA models for pattern formation. In Sections 7.2.3 and 7.2.4 we introduce mathematical treatments of chemical diffusion and reaction. Section 7.2.5 describes pattern formation in reaction-diffusion systems. Section 7.2.6 discusses the coupling of a patterned chemical to additional chemical processes. Finally, Section 7.2.7 describes patterns that might form in vertebrates during development by diffusion of pigment cells from their origin along the spinal cord. A discussion of the relative benefits of CA and reaction-diffusion approaches is included later, in Section 7.4.

As will become apparent in the following sections, creating an interacting system that evolves to a pattern requires us to specify interactions that satisfy various constraints. Since systems evolve toward equilibrium, the principle issues are not dynamic, but rather revolve around constructing a model with a complex pattern as its equilibrium or steady-state structure. In simple systems the equilibrium is homogeneous and has no distinguishable or controllable features. The ability to make patterns requires the specification of a system that behaves in an unconventional manner in equilibrium or steady state.

7.2.2 *Activation and inhibition in pattern formation: CA models*

We begin by thinking about the equilibrium behavior of some simple models. For a CA, the equilibrium is generally described by a stochastic field such as the Ising model

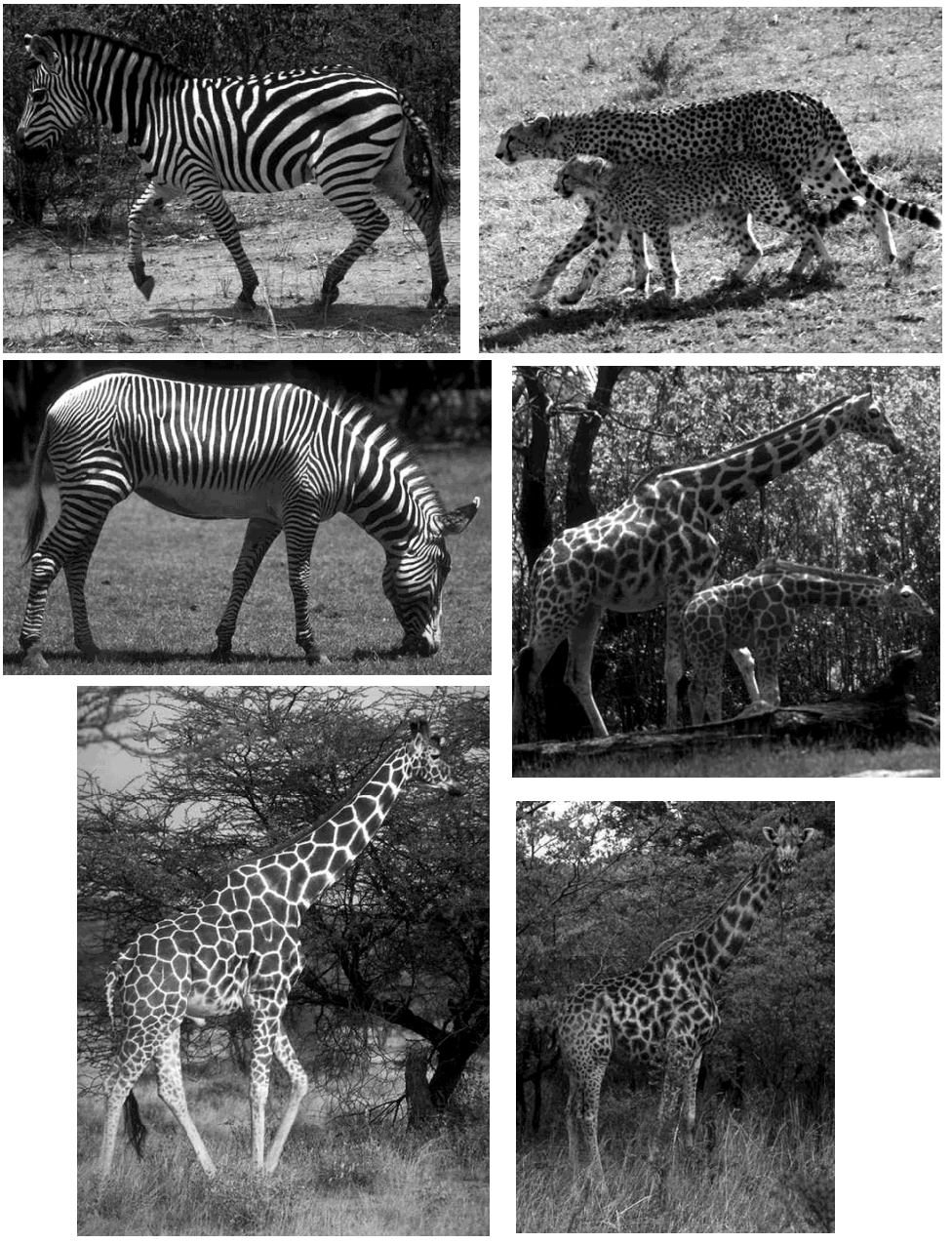


Figure 7.2.1 Photographs showing examples of pigment patterns on animal skins. From top left by row: Grant's zebra, South African cheetah, Grevy's zebra, Uganda giraffe, reticulated giraffe and Masai giraffe. These patterns arise from a process that requires differentiation between regions that contain pigment-producing cells and those that do not. The study of such patterns captures one of the essential processes involved in various stages of development that require differentiation in order to form structures and organs that form a functioning physiology. ∎

(Section 1.6). Since the developmental process leads to a long-lived pattern that remains as the color of the animal,this seems a reasonable starting point. Are there indications that such models can give rise to patterns? The seeds of pattern formation are present in the behavior of an antiferromagnet on a square lattice (Fig. 1.6.7) with alternating values of the variables s_i in its equilibrium state. This pattern arises from simple interactions between neighbors that compel adjacent cells to have opposite values of the spin variable. Considered as a color pattern, it is a checkerboard—the simplest of two color patterns (there are only two such patterns). Is there a way to generalize this to form more elaborate patterns characteristic of animal colors? The most basic feature of the color patterns of animals that is not captured by the checkerboard is the existence of a new length scale. This length scale, the size of dots or bands of color, is characteristic of the pattern. It is not given by the size of the cells or by the size of the animal but rather is a characteristic length scale of its own. It is important to consider how such a length scale can arise. An alternating black and white pattern on the scale of individual cells would appear gray on the scale of the organism.

A straightforward method for creating a new length scale in CA is to extend the range of the interactions between cells. We will take this approach and investigate the consequences. Before we do this let us consider what this means from the point of view of biological cells. It might seem that biological cells interact only with adjacent cells. This interaction occurs by emitting chemicals into the intercellular fluid. The chemicals are then detected by the adjacent cells. Such interactions,however, are not necessarily local, because the distance over which the chemicals travel is controlled by their diffusion constant and lifetime in the intercellular fluid or, more correctly, in the matrix of cells and intercellular fluid. Thus an individual cell can interact with a region of cells in its vicinity, where the size of this region is controlled by the diffusion constant of the chemical as well as reactions that might affect it. More direct modeling of diffusion is discussed in the following section. Here we consider only the effective interaction that results between cells.

In order to generate patterns that consist of a large number of cells that are either all black or all white in regions of a characteristic size, we use interactions that extend a distance typical of the linear dimension of the regions. There are two possible types of pairwise interactions between cells. When a cell producing pigment causes other cells to produce pigment we say that the interaction is activating. When a cell causes others not to produce pigment we say that it is inhibiting. As with the discussion of nerve cell interactions in Chapter 2, the terminology and mathematics of activation and inhibition is similar to the use of ferromagnetic and antiferromagnetic interactions that cause the spins in an Ising model to align or antialign.Spins that are UP are producing pigment, while spins that are DOWN are not. Loosely speaking, a ferromagnetic interaction corresponds to mutual activation. An antiferromagnetic interaction corresponds to inhibition.

How can we design an Ising type model that will give rise to domains of locally aligned spins (either ON or OFF) but will have large scale variation so that adjacent to a region of ON cells there will be a region of OFF cells? The interactions must achieve two effects.First,they must cause the cells that are nearby to have a bias toward having

the same color so that the regions of color are formed. Second,they must have the effect of causing regions that are farther away to have the opposite color. This suggests a short-range interaction that is mutually activating and a long-range inhibiting interaction or, in magnetic language, a short-range ferromagnetic interaction and a long-range antiferromagnetic interaction. This is the model we will be using to obtain various pigment patterns.

It turns out that the magnetic analogy is not without practical application. Real magnetic materials form magnetic domains. The reason for these magnetic domains is that the short-range ferromagnetic interaction between spins is a local effect of quantum mechanics. However, the long-range interaction between spins is through the magnetic field that tries to antialign the spins—an antiferromagnetic interaction. This gives rise to domains of magnetization that form a pattern of regions of UP and DOWN spins that has a large scale compared to the atomic distances. When a piece of iron is magnetized,it is forced into a metastable state by aligning these magnetic domains. After long enough time,it demagnetizes by returning to its equilibrium state. Modern use of patterns of magnetization appears in magnetic bubble memories that vary external fields to manipulate the patterns of magnetic domains very much in the manner described below.

We will adopt the Ising model terminology of spin variables to describe pattern formation. In Fig. 7.2.2 the spin-spin interaction for a model of pattern formation is plotted as a function of distance. The energy of the system would be written as:

$$E[\{s_i\}] = -h \sum_i s_i - \frac{1}{2} \sum_{i,j} J(r_{ij}) s_i s_j \tag{7.2.1}$$

where $s_i = \pm 1$ is ON and OFF respectively. $J(r_{ij})$ is the interaction as a function of distance r_{ij} between spins. This is similar to Eq.(1.6.52) but includes only a uniform bias field h that controls how likely a cell is to have pigment (ON) as opposed to no pigment (OFF). Writing explicitly the interaction in terms of two parameters $J_1 > 0$ and $J_2 < 0$ we have:

$$E[\{s_i\}] = -h \sum_i s_i - \frac{1}{2} J_1 \sum_{r_{ij} < R_1} s_i s_j - \frac{1}{2} J_2 \sum_{R_1 < r_{ij} < R_2} s_i s_j \tag{7.2.2}$$

What should we expect from the equilibrium structure of this model? The main point is that the existence of long-range antiferromagnetic interactions should cause patches of color. However, we have already found in some cases that the presence of antiferromagnetic interactions causes many low-energy states rather than only a single unique one. While this was true in Section 1.6 only for nonbipartite lattices, we anticipate that it will be true for this more complicated model. Thus we will avoid trying to describe directly the equilibrium states of this model and focus instead on what we are more interested in anyway—the outcome of its dynamics. For convenience, we take a square lattice and start from a random set of values with half of the cells ON. We construct a dynamics for the system, then run it until there are no changes and record the resulting pattern.

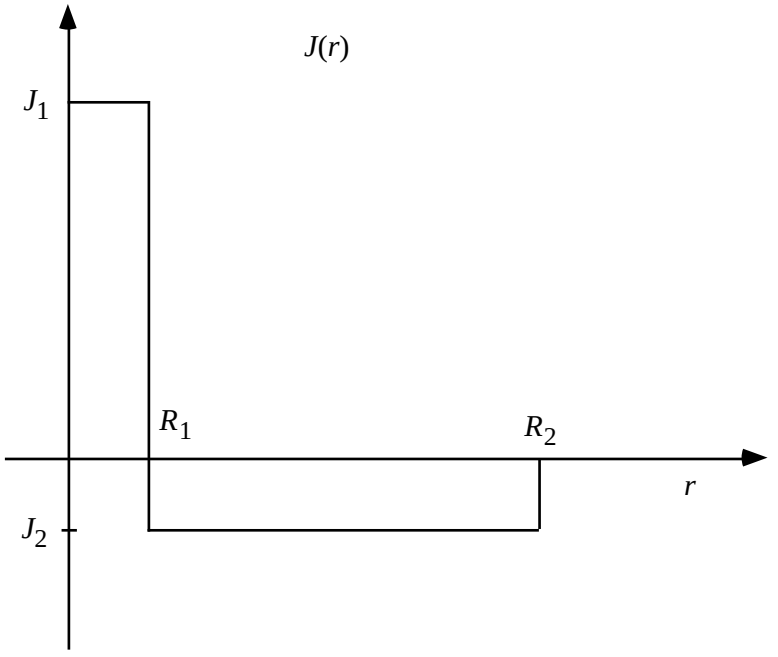


Figure 7.2.2 A CA model of pattern formation uses interactions that cause short-range activation and long-range inhibition of pigment. The interaction as a function of distance $J(r)$ in this model is illustrated. The same model describes interactions that are locally ferromagnetic and long-range antiferromagnetic in a magnetic system. ∎

The dynamics is the same as that used in Chapter 2 for the neural network—zero-temperature Glauber or Monte Carlo dynamics. We choose a particular cell to update and set it ON or OFF according to which gives the lower energy. Stated differently, the cell is set ON if the net interaction of the cells causes it to be ON. The total influence of the other cells is given by the effective field:

$$h_i = h + J_1 \sum_{r_{ij} < R_1} s_j + J_2 \sum_{R_1 < r_{ij} < R_2} s_j \tag{7.2.3}$$

We thus set the value of $s_i(t)$ to be:

$$s_i(t) = \text{sign}\left(h + J_1 \sum_{r_{ij} < R_1} s_j(t-1) + J_2 \sum_{R_1 < r_{ij} < R_2} s_j(t-1)\right) \tag{7.2.4}$$

This equation is quite similar to the equation describing the update of neural cells Eq.(2.2.4). The difference between Eq.(7.2.4) and Eq.(2.2.4) is how we set the values of the interactions between the spin variables, and the presence of a bias h. The pigment cells are locally interacting, while in Chapter 2,neural cells were interconnected

throughout by interactions J_{ij}. In Chapter 2 we considered the update of cells to be synchronous because in the presence of random interactions this does not generally cause different results. Here, it is better to update the system asynchronously by selecting cells to update sequentially at random. This avoids oscillations that can occur when all the cells are updated simultaneously.

There are five parameters in this model:the two interaction ranges R_1 and R_2, the two interaction strengths J_1 and J_2 and the bias field h. Since we have not yet chosen the scale of the interaction strength, we can choose it so that one of them takes a convenient value. We set $J_1 = 1$. It is positive,as required for a ferromagnetic interaction. J_2 takes a negative value. It makes sense to choose a value of J_2 smaller than J_1 because J_2 acts over a larger area. We choose $J_2 = -0.1$. We set the value of R_1, the range of the short-range interaction, to a nearest-neighbor distance or $R_1 = 1$. Distance is measured in the cellular space by cell size. The range of R_2 should have something to do with the size of the pattern elements that result. We set this to a value of $R_2 = 6$ to have a large enough value that will be distinct from the nearest-neighbor distance and small enough to be comfortably within the space we simulate, which will be 60×60 cells. We start by setting the value of the bias field $h = 0$ and vary it to create patterns with more or less ON or OFF cells.Fig. 7.2.3 illustrates the generation of a pattern from a random starting configuration of the cells. The computer program used to generate these patterns is similar to those used in Sections 1.5 and 1.6 to investigate the dynamics of cellular automata and the Ising model respectively. We can see that long-range inhibition gives rise to alternating regions of colors at a characteristic separation distance.

Fig. 7.2.4 illustrates the variation in patterns that can be generated as a result of changing the value of the bias field h. All of the patterns on the left result from the same initial random array of dots. Similarly, all of the patterns on the right result from a single but different random initial array of dots. Considering the left patterns and the right patterns separately, we can see how the change in the bias field affects the eventual pattern that is formed. At one extreme there are black dots in a sea of white. The dots are not regularly spaced or shaped. They are variable in size and some are elongated. As the value of h is increased, more dots elongate and connect forming bands that interconnect and eventually become the black background in which white dots exist. These patterns are reminiscent of some animal color patterns.

In Fig. 7.2.5 we investigate the effect of increasing the value of R_1 (right panels) and R_2 (left panels). The most obvious changes occur with R_2. The characteristic size of the pattern increases and is directly controlled by R_2. Increasing R_1 does not affect the size of the pattern but rather the shape of the boundaries between regions of ON and OFF cells. Increasing R_1 ensures that the boundaries of dots and stripes are smoother, with more gradual changes in curvature. This is particularly apparent in our simulations because the size of R_1 is comparable to the size of cells. For more realistic animal color patterns, the size of R_1 should be larger than the size of cells, to avoid sharp corners.

The initial conditions of the simulation can be important. We started these simulations with 50% of the cells set ON at random. The effect of the initial random

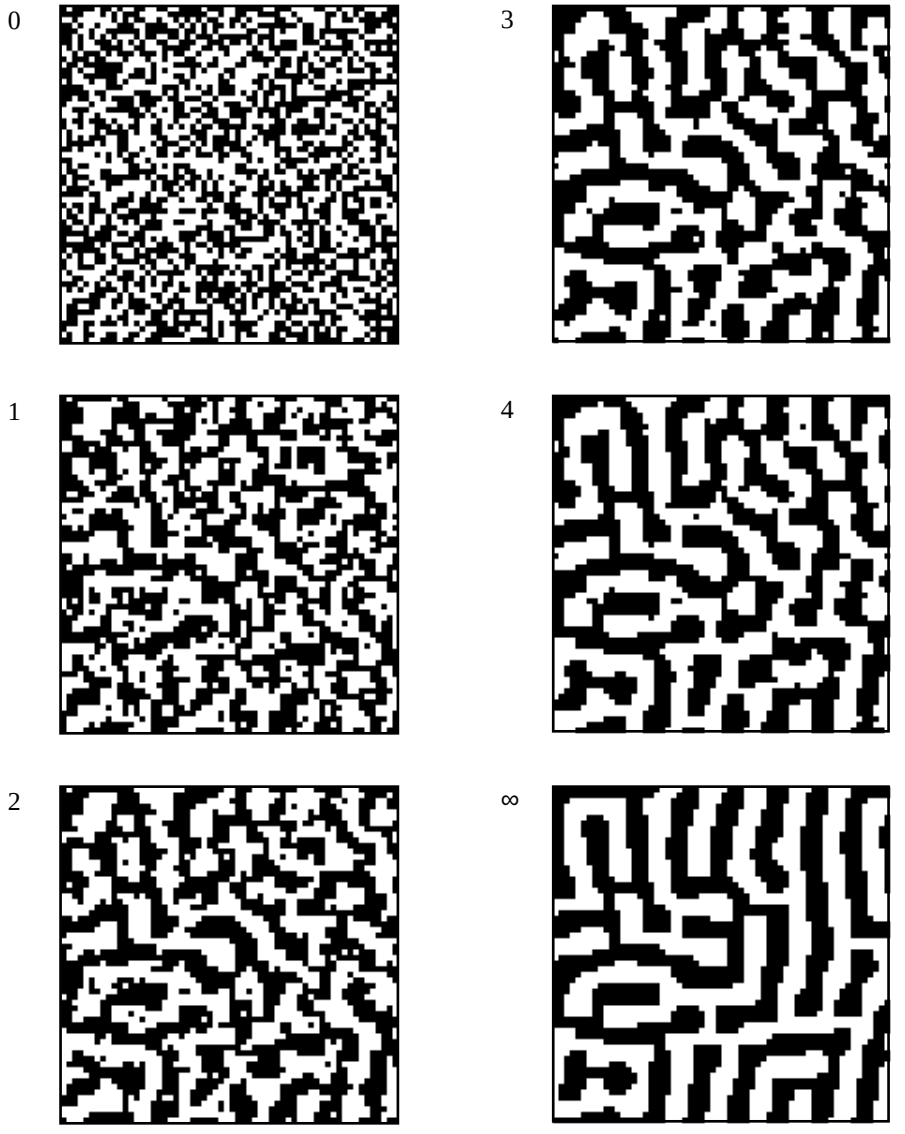


Figure 7.2.3 A simulation of a CA model of pattern formation. ON cells (black) produce pigment and OFF cells (white) do not. The initial conditions assign cells to be ON or OFF at random with probability 1/2. Five updates are shown and then the unchanging stable limit that is reached after about 20 updates. The parameters are $R_1 = 1$, $R_2 = 6$, $J_1 = 1$, $J_2 = -0.1$, and $h = 0$. ∎

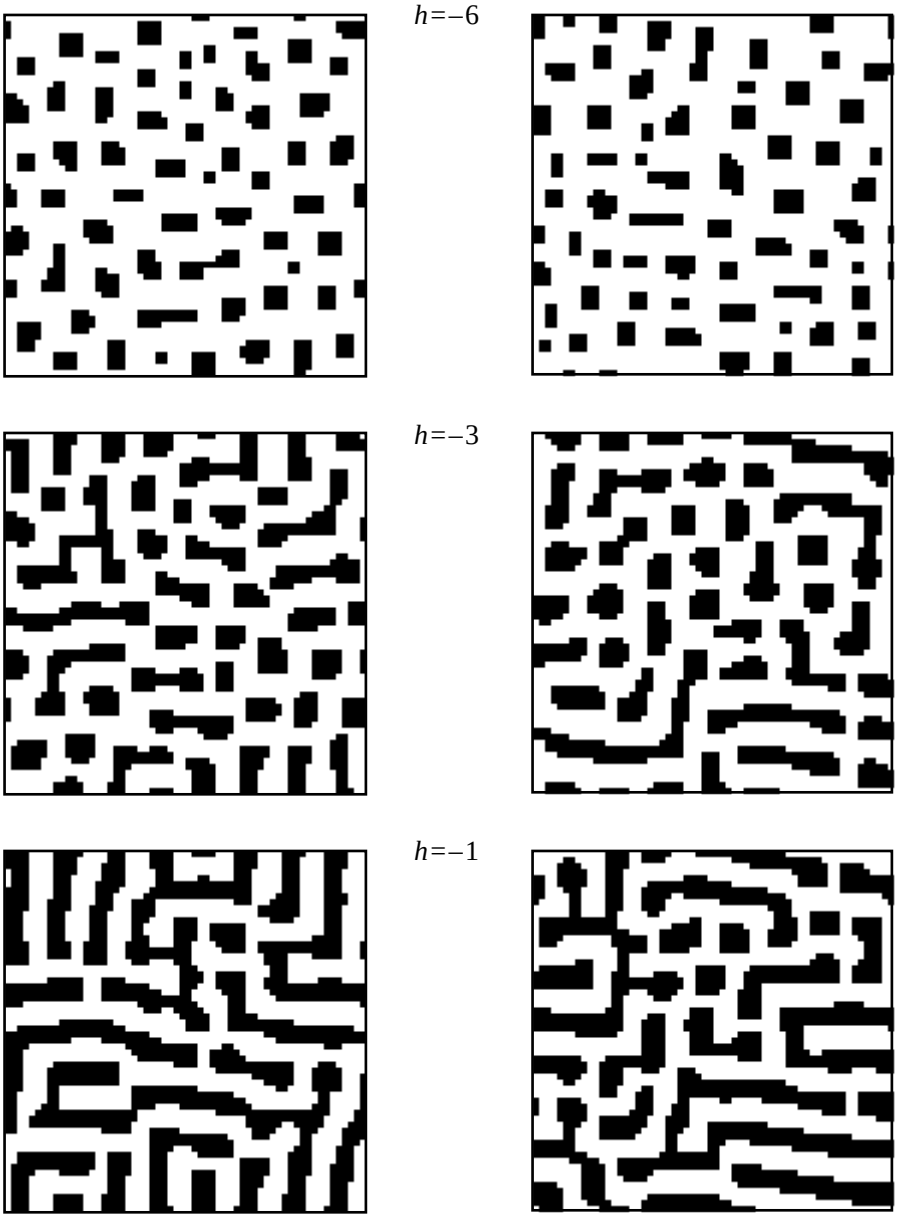
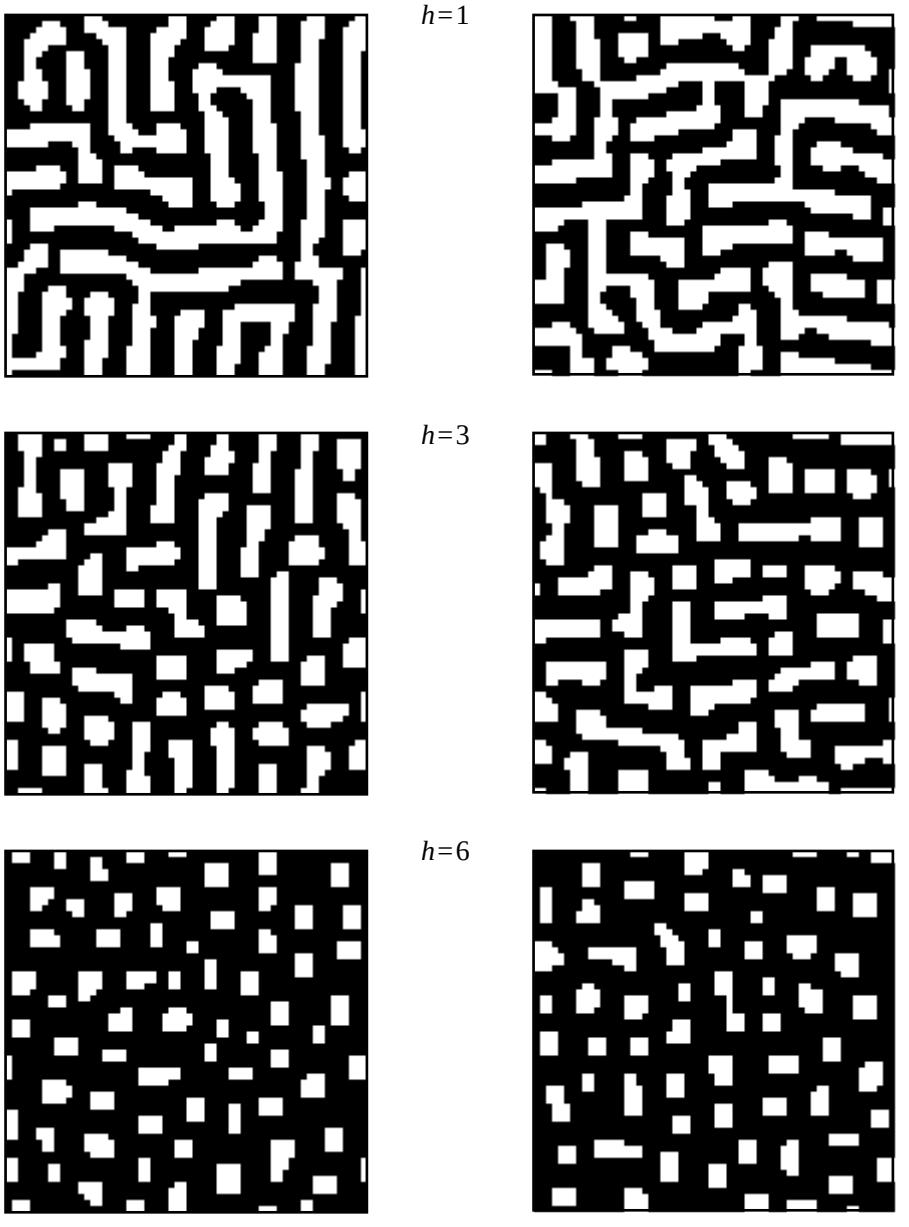


Figure 7.2.4 Additional simulations of the CA model that illustrate the effect of varying the bias field *h*; other parameters are the same as Fig. 7.2.3. All patterns shown are the unchanging stable limit of a simulation. *h* biases the system to have more or less ON cells. Varying *h* results in patterns with black spots on a white background, white and black stripes, or white spots on a black background. The left and right panels differ only in the initial conditions of the simulation. All of the left panels start with the initial condition shown



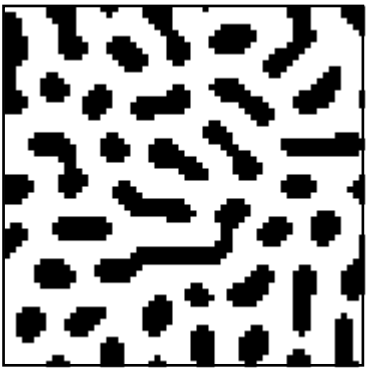
in Fig. 7.2.3. The right panels begin from a different random initial condition. We see that left and right panels share general characteristics but are different in detail. While both initial conditions have a probability of 1/2 that cells are ON, qualitative aspects of the final patterns are not sensitive to the initial probability, since they are determined by the ensemble of stable states of the system. ∎

R_2=6.0 R_1=1.0 h=0

R_2=6.0 R_1=1.5 h=0

R_2=7.0 R_1=1.0 h=0

R_2=6.0 R_1=1.5 h=–3



R_2=8.0 R_1=1.0 h=0

R_2=6.0 R_1=1.5 h=–6

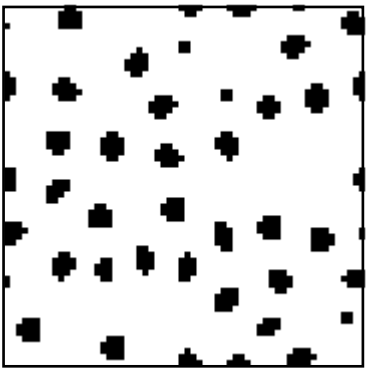


Figure 7.2.5 Changing other parameters in the CA model. Each pattern is the steady state of a simulation with the parameters indicated, and for all cases $J_1 = 1$ and $J_2 = -0.1$. The left panels show the variation in the spatial scale of the pattern that results from changing the range R_2 of the antiferromagnetic interaction. Simulations for three different values of R_2 are shown. There is a direct relationship between R_2 and the size of the stripes. The right panels show patterns that arise when the range R_1 of the ferromagnetic interaction is increased. Simulations with three values of h are shown all with the same increased value of R_1. The top right panel should be compared with the top left panel. The other two right panels should be compared to the panels of Fig. 7.2.4 with the same value of h. The effect of the increase in R_1 is to round the corners of the spots and stripes. ∎

configuration is apparent from the nonuniform nature of the pattern,and the two different results shown in left and right panels of Fig. 7.2.4. Changing the initial proportion of ON cells has very little effect on the qualitative behavior of the model because the resulting pattern is essentially an equilibrium pattern—one of many with similar number and shapes of color regions. However, the specific pattern of dots and their shapes is sensitive to the precise starting pattern of ON and OFF cells. If we consider this as a theory of the origin of animal color patterns, it suggests that individual differences may be due to randomness rather than genetic control, while the overall characteristics are controlled by the underlying mechanism, which is genetic and species specific. In this case the particular pattern is not heritable,and even identical twins would have different patterns. This should not be the case with many other characteristics.

Question 7.2.1 In the model we have just simulated,patterns appear to arise in equilibrium. We have argued in Section 1.3 that equilibrium systems have simple behavior. Why doesn't this apply in our case?

Solution 7.2.1 The thermodynamic limit discussed in Section 1.3 applies when we take the limit of large enough system size. The results in this section do not apply when we take this limit, since then the system would appear uniform and homogenous, because the size of the spots that we are discussing would be so small as to be irrelevant. When this limit is not used, then the conclusions also do not apply.

A more thorough discussion would note that there are actually two conditions that are not met by these systems, consistent with the discussion in Section 1.3.6.

First,the ergodic theorem does not apply. This means that the ensemble of possible states of the system is not being explored. This is apparent when we consider that the system iterates to a steady state and that this steady state is a unique state that is unchanging even though there are many such possible states. Moreover, when the ergodic theorem applies,the initial state is irrelevant to the final equilibrium state. The reason that this model breaks the ergodic theorem in such a direct way is that we are modeling it at zero

temperature—the temperature is so low that no random changes occur. The only changes are those dictated by energy reduction. In a system where temperature causes random changes, there would be a time dependence to the pattern. If our observations of such a system were averaged over a long time, we would not see any specific pattern, but only a homogeneous average. If our observations were averaged over a shorter time, we would see an individual pattern.

Second,in these simulations a correlation length exists that is not small on the scale of the whole system. The length scale that is relevant is the characteristic length scale of the pattern. In some patterns it may actually be larger than the size of the stripes or dots,since the positions of stripes can be correlated with each other. We can see the relevance of the pattern length scale by considering what would happen if we observed a system that was much larger than this length scale. Then the pattern would become irrelevant and the color would be gray on the scale of our observations.

The relevance of temporal and spatial scale to the complexity of a system will be discussed further in Chapter 8. ▮

Question 7.2.2 Consider a model using variables $\bar{s}_i = 0,1$ to represent unpigmented and pigmented cells. We will use overbars to indicate all quantities in this model. Set the update rule to be similar to that in Eq.7.2.4 but with the bias field $\bar{h} = 0$. When the effective local field $\bar{h}_i$ is negative, the cell is set to 0; when it is positive the cell is set to 1. This is a more intuitive representation of activation and inhibition since both are effected by cells that are ON. Cells that are OFF have no influence on the pigment production in other cells. It is also assumed that there is no tendency for cells to spontaneously become pigment producing. Simulate this model and vary the strength of the inhibition $\bar{J}_2$ to obtain various patterns. How can this model be transformed back to that given in the text?

Solution 7.2.2 Using the same parameters as the text except for the value of J_2 and h, the results of simulations are shown in Fig. 7.2.6.

To transform this model to that in the text, we can perform the substitution $\bar{s}_i \to (s_i + 1)/2$ so that $0,1 \to -1, +1$.Once this substitution is performed in Eq.(7.2.3) we can recognize the parameters that would give the same patterns in the original model:

$$\begin{aligned} J_1 &= \bar{J}_1/2 \\ J_2 &= \bar{J}_2/2 \\ h &= \bar{J}_1/2 \sum_{r_{ij}<R_1} 1 + \bar{J}_2/2 \sum_{R_1<r_{ij}<R_2} 1 \end{aligned} \qquad (7.2.5)$$ ▮

Question 7.2.3 Consider what will happen if $R_1 = 0$, i.e., there is no activation in the model of the text and $\bar{R}_1 = 0$ for the model of Question 7.2.2.

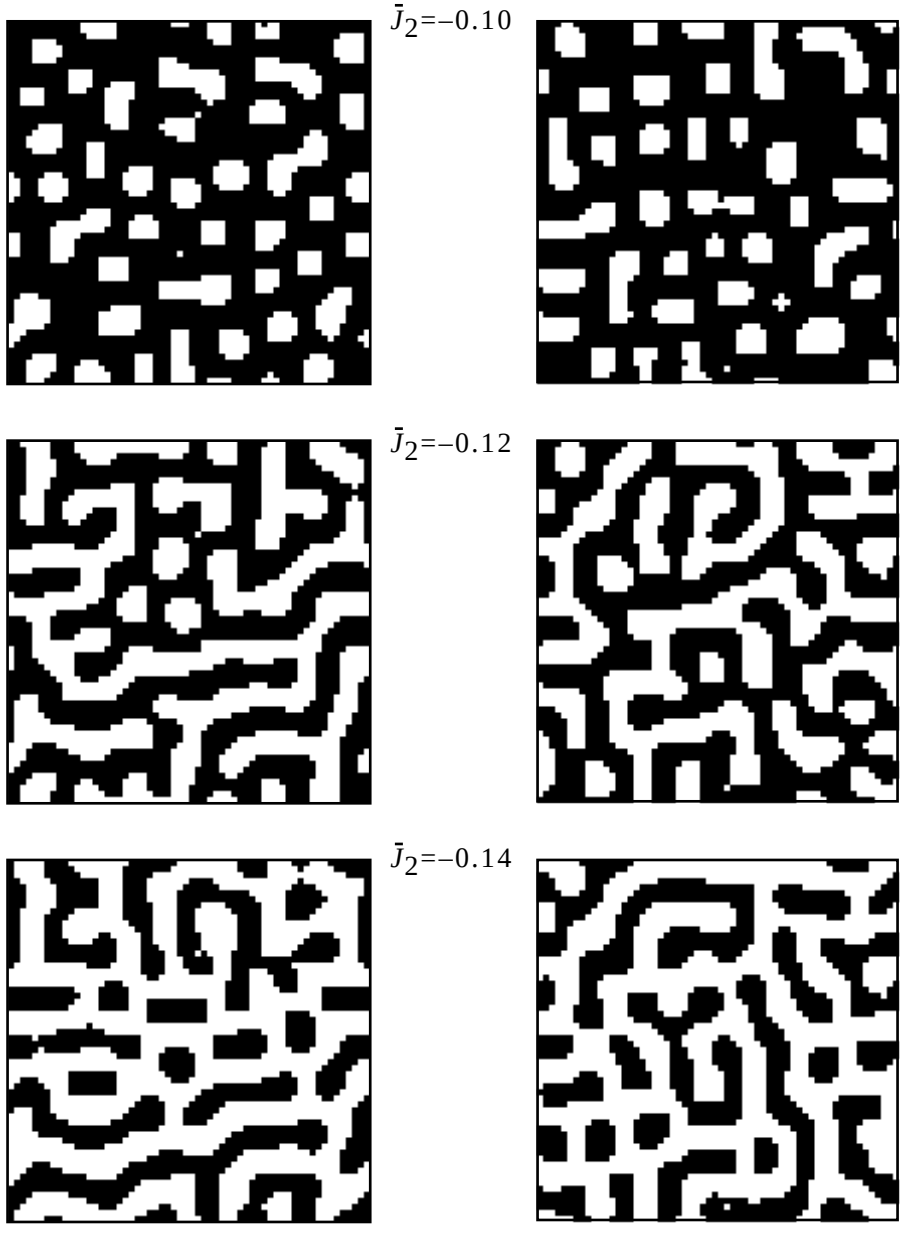


Figure 7.2.6 Using a different parametrization of the CA model for pattern formation with $\bar{s}_i = 0, 1$ and $\bar{h} = 0$ in Eq. 7.2.3, we generate patterns that are similar to Fig. 7.2.3 by varying the strength of the inhibition $\bar{J}_2$. The other parameters were taken to be $\bar{R}_1 = 1$, $\bar{R}_2 = 6$, $\bar{J}_1 = 1$. Left and right panels use different initial random conditions similar to Fig. 7.2.4 (see Question 7.2.2). ∎

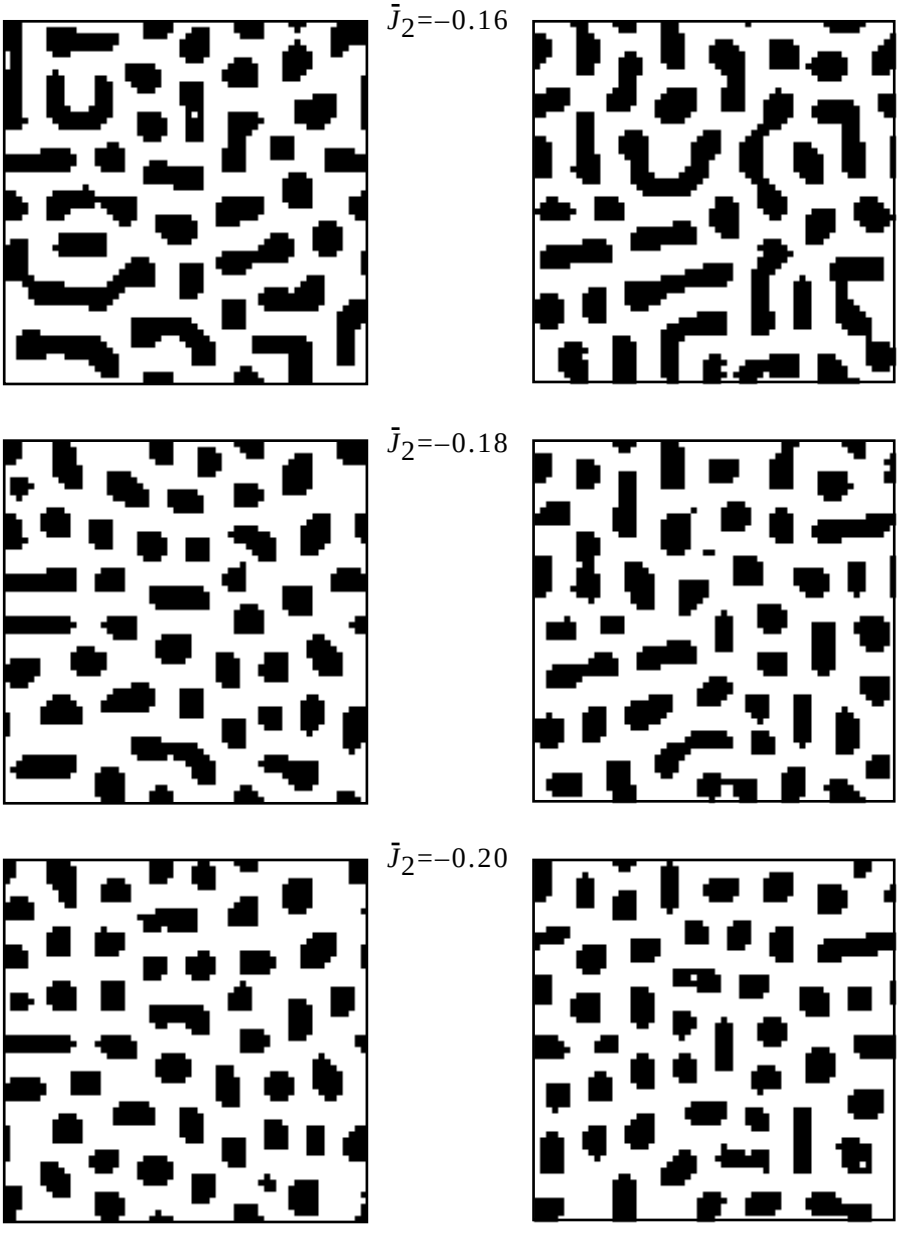


Figure 7.2.6 (*continued*)

Solution 7.2.3 Since the pattern formation seems to depend on local activation and long-range inhibition, we might think that setting $R_1 = 0$ would eliminate pattern formation. However, in the case of the model discussed in the text, patterns still form. In order for the long-range antiferromagnetic interaction to lower the system energy, it is necessary for regions to be locally ferromagnetic. Another way to understand this is that the long-range interaction controls the long-range properties of the pattern while, as seen in Fig. 7.2.5, only the boundary shapes are controlled by the short-range interaction. We could even have a short-range antiferromagnetic interaction and still have patterns, as long as the short-range interaction is not too strong. However, when we consider the model of Question 7.2.2 we realize that in order for cells to turn ON there must be a local activation, otherwise cells can only turn OFF in the dynamics. From Eq.(7.2.5) because $\bar{J}_2 < 0$, $\bar{J}_1 = 0$ would correspond to a large negative h in the model discussed in the text. The large negative h would likewise prevent any cell from turning ON. ∎

The patterns that can be generated using the activation-inhibition CA model suggest that the variability between different species can be readily achieved by variations in parameters of such models. However, this particular set of patterns does not capture the appearance of many of the common animals. One example is the giraffe (specifically the Uganda giraffe), which has patterns of coloration characterized by regions of pigment separated by relatively narrow and straight lines without pigment. We will discuss an approach to generating such patterns which illustrates that there may be other mechanisms for generating patterns of a certain scale. The method begins by noting (Fig. 7.2.7) that giraffe patterns appear to be similar to patterns generated in two steps. First we choose a sparse set of initial dots. Then we divide the plane into regions associated with each dot. The region associated with a dot consists of all points that are closer to it than any other dot. Then the boundaries of these regions are not colored, while the interiors are.

To generate this pattern, we use a CA that grows regions of color from isolated points, which are cells initialized ON. The growing regions then stop when they reach the proximity of another region that is growing. In this rule, the characteristic size of the pattern is given by the density of the initial ON cells. This would be similar to a process of nucleation and growth (Section 1.6.8), where nucleation creates the isolated points that expand rapidly compared to the nucleation time. The CA rule we use is similar to those described in Sections 1.5.2 and 1.5.3, for the condensation model and Conway's Game of Life. We will construct the rule step by step.

To allow regions to grow, a cell is set ON at time t if at time $t - 1$ there were more than zero cells ON in its neighborhood. This results in growth from a point expanding into the space in a uniform fashion. Because of our square lattice, there is a problem in the shape of growth—it is not circular as would be expected in a physiological system. By expanding the range of influence of a single cell, which corresponds to increasing the size of the neighborhood, we can make it more circular, as illustrated in Fig. 7.2.8.

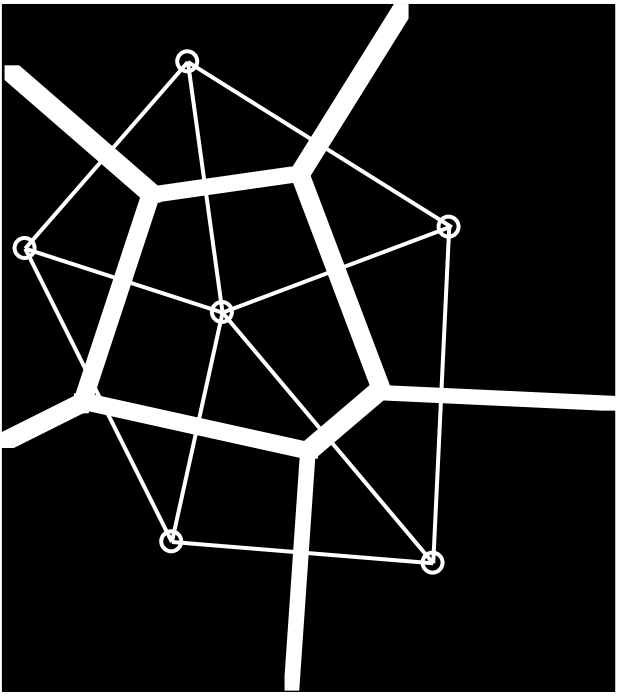


Figure 7.2.7 The patterns of coloration on a giraffe can be understood geometrically. They appear to be generated by dividing the two-dimensional surface according to their distance from a sparse selected set of points. In this figure the selected set of points are indicated by circles. Line segments connecting them are shown as thin lines. By coloring areas that are close to each of the selected points, but not points that are approximately the same distance from two or more points, we can generate patterns similar to those found on some species of giraffe. ∎

In order to leave uncolored the regions between growing dots, cells must recognize when the two growing regions meet. When the pigment grows from a point, the shape of the ON region is convex. We can identify a cell in the encounter region because it has more ON cells around it than a cell at the boundary of the growing region. A cell that has more than a certain number of neighbors with pigment must be in the encounter region and should not turn ON. Thus, the CA sets a cell ON if it has some but not too many ON neighbors. Note that in this model we are not allowing cells that are ON to turn OFF. This is important, because otherwise cells in the interior of a spot would turn OFF once we impose the condition that stops the growth.

We start the growth by setting cells ON at random with a probability of 1 in 100. The result of this simulation,illustrated in Fig. 7.2.9, is not very satisfactory. Some of the cells at the boundaries of growing regions do not turn ON. However, these do not form continuous lines. To overcome this problem we need to have wider regions of

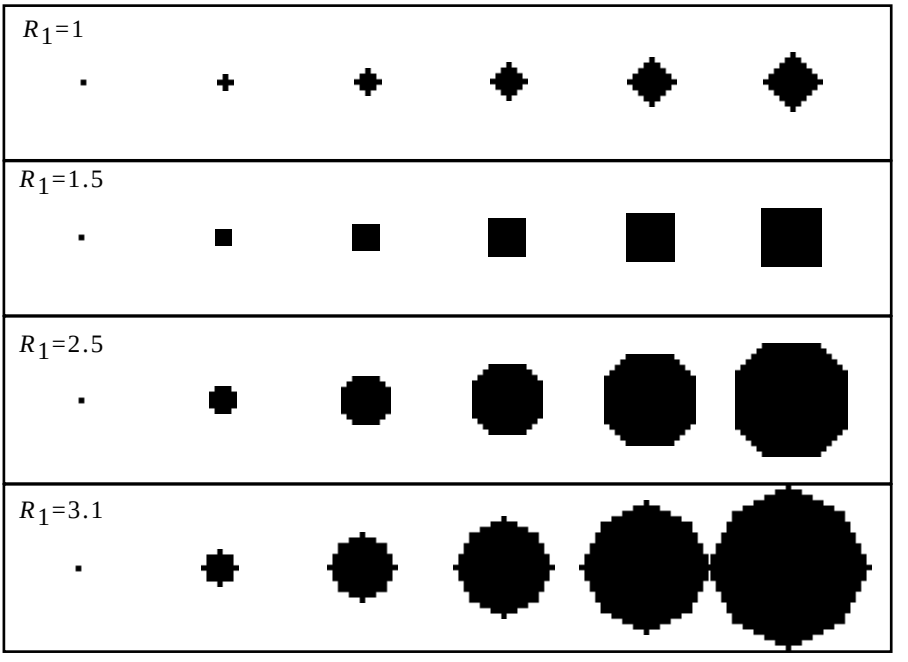


Figure 7.2.8 Starting from a seed pigment cell we can grow outward using a rule that sets a cell ON if there are any ON cells in its neighborhood. However, the shape of the growing region on a square lattice depends on the way we grow it. Here, growth of a region is shown for various sizes of the neighborhood given by its radius R_1. A larger R_1 leads to more circular pigmented areas. ∎

OFF cells. To achieve the desired result, we can take a clue from the previous model of pattern formation and set up two distances,a distance R_1 over which the growth is determined, and a distance R_2 over which the stopping is determined. Thus, using binary variables $s_i = 0,1$ we turn a cell ON when the value of Eq.(7.2.3) is positive, with parameter values of $R_1 = 2.5$, $R_2 = 4.3$, $J_1 = 1$ and $J_2 = -.5$. The values of these parameters can be adjusted by trial and error.

The patterns generated in Fig. 7.2.10 using this approach are reminiscent of the patterns of giraffes;however, they are not entirely satisfactory. While some of the regions follow the convex shape that we expect,other regions are more convoluted. By looking carefully at the patterns, we see that this occurs because the separations between the initial ON cells vary in distance. This would not occur if the starting points were more regularly spaced. There are many ways to consider placing the points at more regular intervals. A reasonable approach for this case is to use the previous method of creating patterns using activation-inhibition to generate a pattern of spots such as those shown in Fig. 7.2.4 and then to apply the growth starting from these spots. This is illustrated in Fig. 7.2.11, where the initial pattern is generated

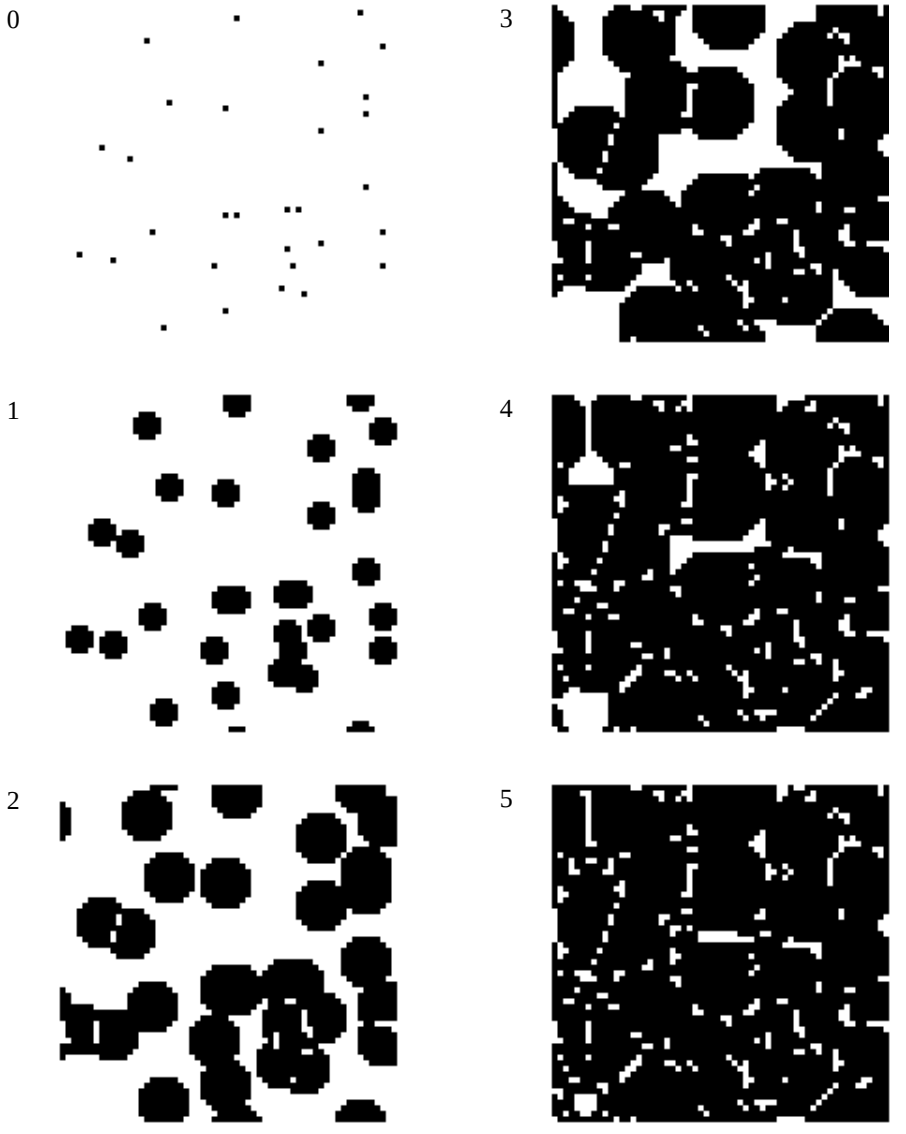


Figure 7.2.9 A first attempt at forming a color pattern similar to that of a giraffe. The initial conditions are obtained by setting cells to be ON at random with a small probability, here taken to be 1 in 100. The algorithm updates the cells synchronously and sets them ON if the number of cells in a neighborhood of radius $R_1 = 2.5$ is nonzero, but also less than 10. The color grows out from the initial ON cells. When growing regions meet, there are some cells that do not turn ON because of the limiting condition on the number of cells in the neighborhood. However, these regions of residual OFF cells are not continuous. ∎

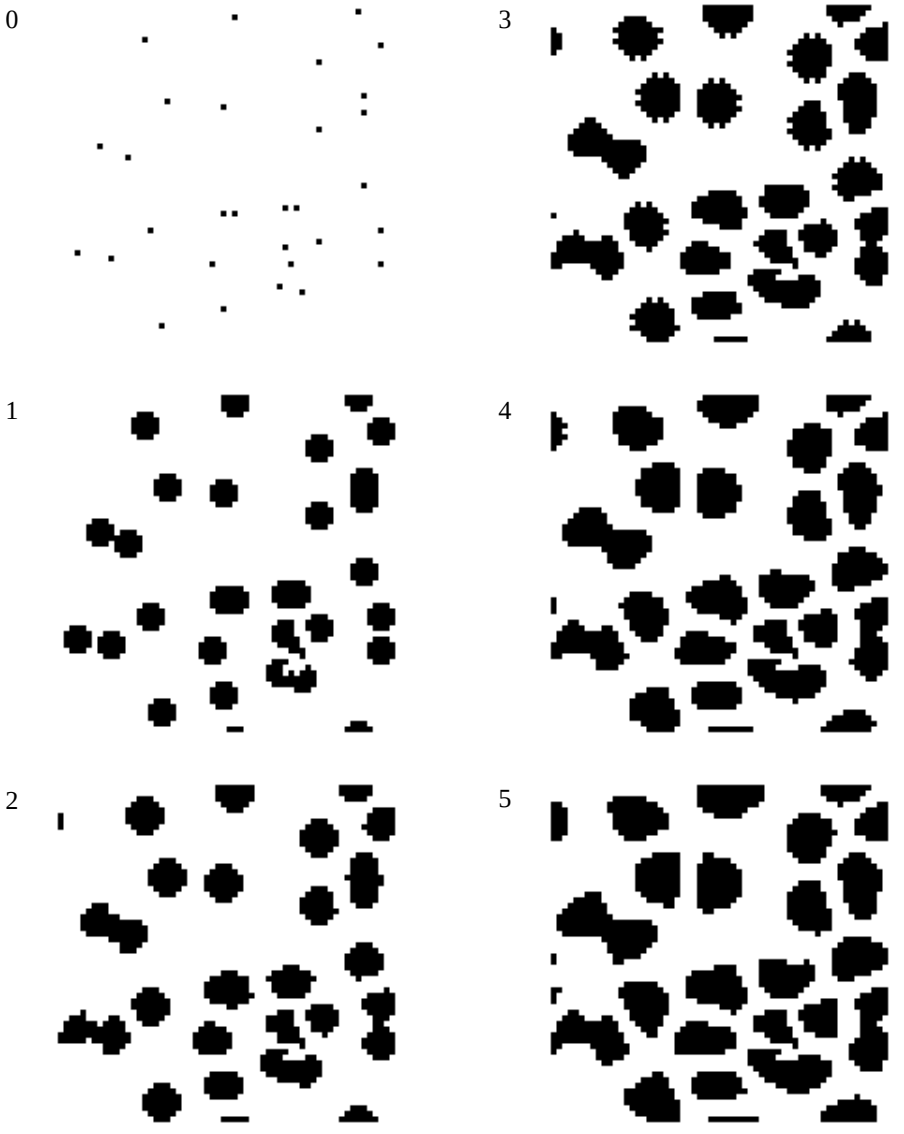


Figure 7.2.10 Better simulations of the formation of giraffe patterns than those in Fig. 7.2.9 result if we use a larger region to set the condition for stopping growth. The parameters, adjusted by hand, are inspired by the activation-inhibition model. The growth results from activation of cells adjacent to cells that are ON, while too many ON cells in a larger region (long-range inhibition) cause cells not to turn ON. Here a particular simulation is shown from its initial condition for seven updates, and then the final stable result. Three outcomes starting from other random initial conditions are shown in the rightmost column. All initial conditions were set with a probability 1 in 100 of cells being ON. These simulations are not entirely satisfactory because many of the spots have unusual shapes. ∎

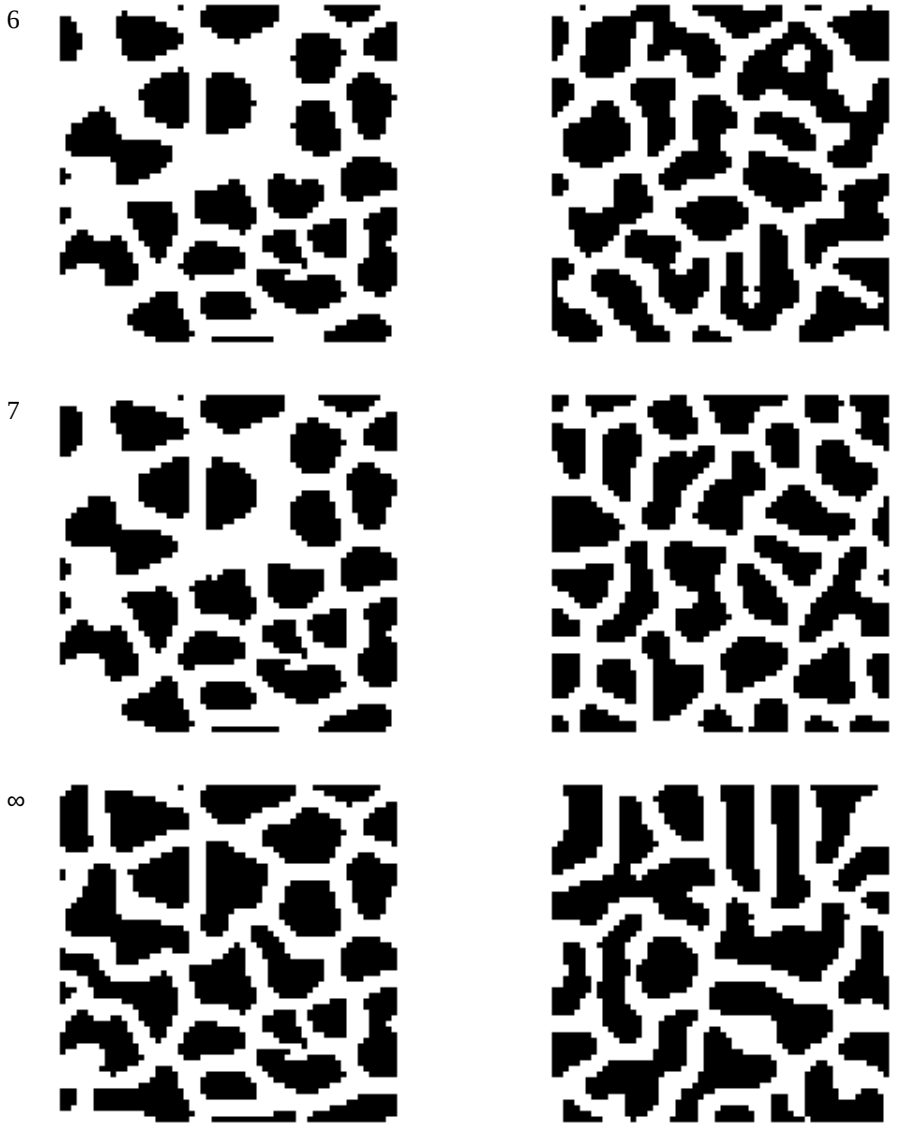


Figure 7.2.10 (*continued*)

from a CA activation-inhibition model, resulting in more regular but still randomly placed spots. By growing out into the OFF regions we form a pattern that is closer to the patterns on the giraffe coats. More specifically, this coloration is similar to that of the Uganda giraffe (Fig. 7.2.1). Two other kinds of giraffe—the reticulated giraffe and the Masai giraffe—would require additional tuning of parameters. The reticu-

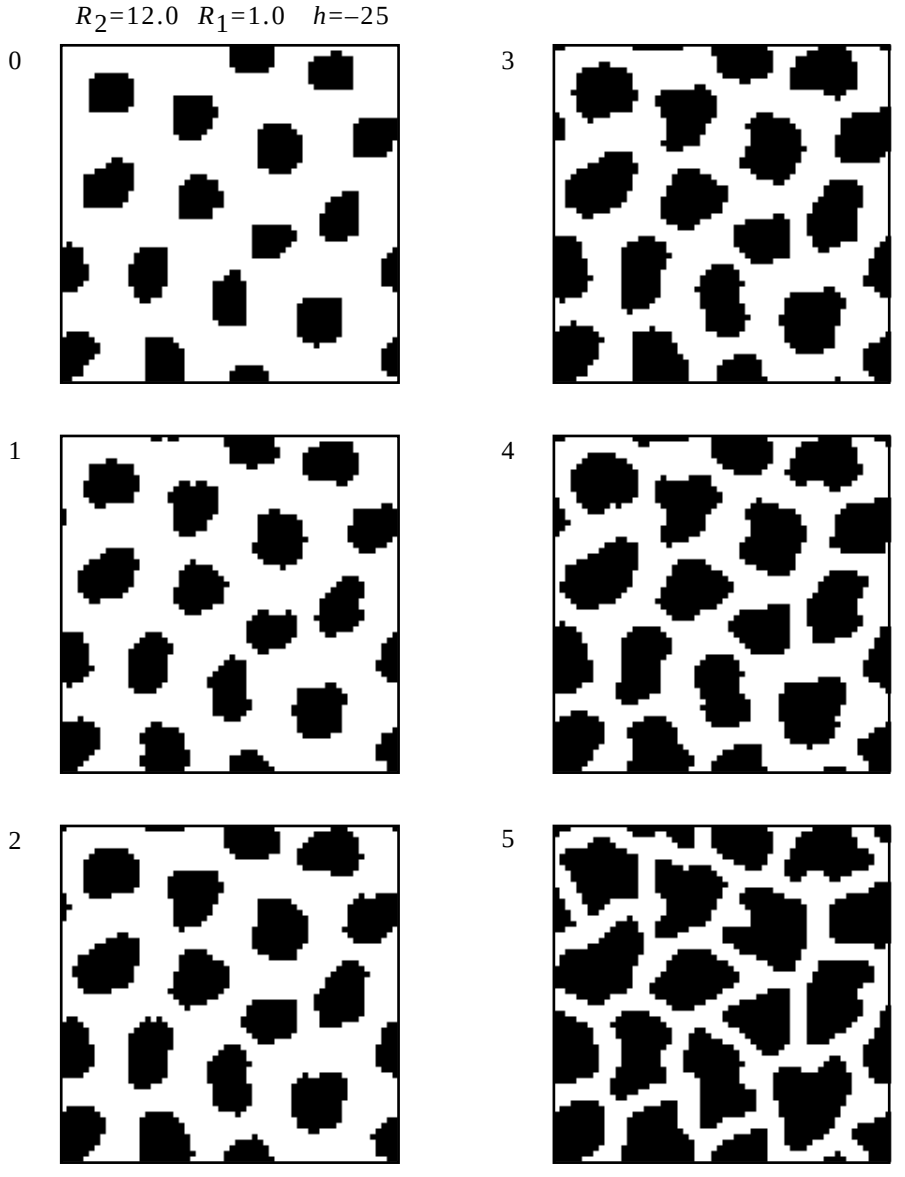


Figure 7.2.11 The giraffe color patterns generated in Fig. 7.2.10 can be improved by starting from points that are more regularly spaced in the plane. They might be placed more regularly by several processes, one of which is illustrated here. The initial conditions result from an activation-inhibition CA model simulation with parameters as indicated on the upper left. This is the starting conformation for the growth outward of pigmented regions. The subsequent frames show updates using the same algorithm as Fig. 7.2.10. This results in a more regular pattern reminiscent of the Uganda giraffe. Other patterns can be generated by varying the parameters. ∎

lated giraffe would be generated by a smaller ratio of the line width to the size of the spots. This requires a finer mesh of points but could be simulated by the same algorithm. The third kind of giraffe, the Masai giraffe, has spots that are blotches with fingering. Such fingering can also be achieved by varying the parameters in this algorithm.

7.2.3 *Chemical diffusion*

We can add an additional layer of detail in our models by considering more directly the properties of molecules produced in cells and their motion through the matrix of cells and intercellular fluid. Molecules generally move by a random walk that is not directed but results from the random thermal motion of the liquid, mostly water, in which they are located.A single molecule undergoing a random walk travels a characteristic distance proportional to the square root of the time, or $\sqrt{Dt}$, where D is the diffusion constant. The probability distribution of the behavior of a single molecule also describes what happens to a density of weakly interacting molecules. If there is a localized density of molecules at one place,it will spread over time and the distribution will approximate a Gaussian that broadens and flattens over time (Section 1.2). This molecular motion,diffusion,in the continuum limit is described by a differential equation (Section 1.4.4) that represents the changes in density $n(x;t)$ with time when it is sufficiently smooth:

$$\frac{dn(\mathbf{x};t)}{dt} = D \quad ^2 n(\mathbf{x};t) \tag{7.2.6}$$

This discussion suggests that we consider pattern formation arising from a differential equation representing the evolution of molecular density. This approach was taken by Turing, more generally known for the invention of Turing machines discussed in Section 1.9.4. The resulting color patterns are known as Turing patterns.

The CA approach in the previous section treated diffusion as an incidental process which was summarized by an effective interaction between the cells. This simplified the study of the process of pattern formation so that the activation and inhibition were readily apparent. In this and the following section we construct two essential parts of the differential equation approach—the diffusion and reaction of molecules. Then we discuss and simulate specific sets of equations that give rise to patterns.

We derived the diffusion equation (Eq.(7.2.6)) in Section 1.4.4 from the motion of a particle in a periodic set of wells. It is more usually derived from the motion of a low-density "gas" of molecules that have a varying density profile as a function of position as illustrated in Fig. 7.2.12(a). We consider the current $J(x)$ of molecules at a particular position x and relate this current to the variation of the density with position $n(x)$. In order to obtain the current, we make use of simplifying assumptions. The result is more general than the assumptions suggest. We assume that molecules undergo instantaneous collisions with a fluid or matrix in which they are embedded. The characteristic time between collisions is τ. In between collisions,particles have a characteristic velocity v and travel a distance $l = v\tau$. v is determined by thermal motion—

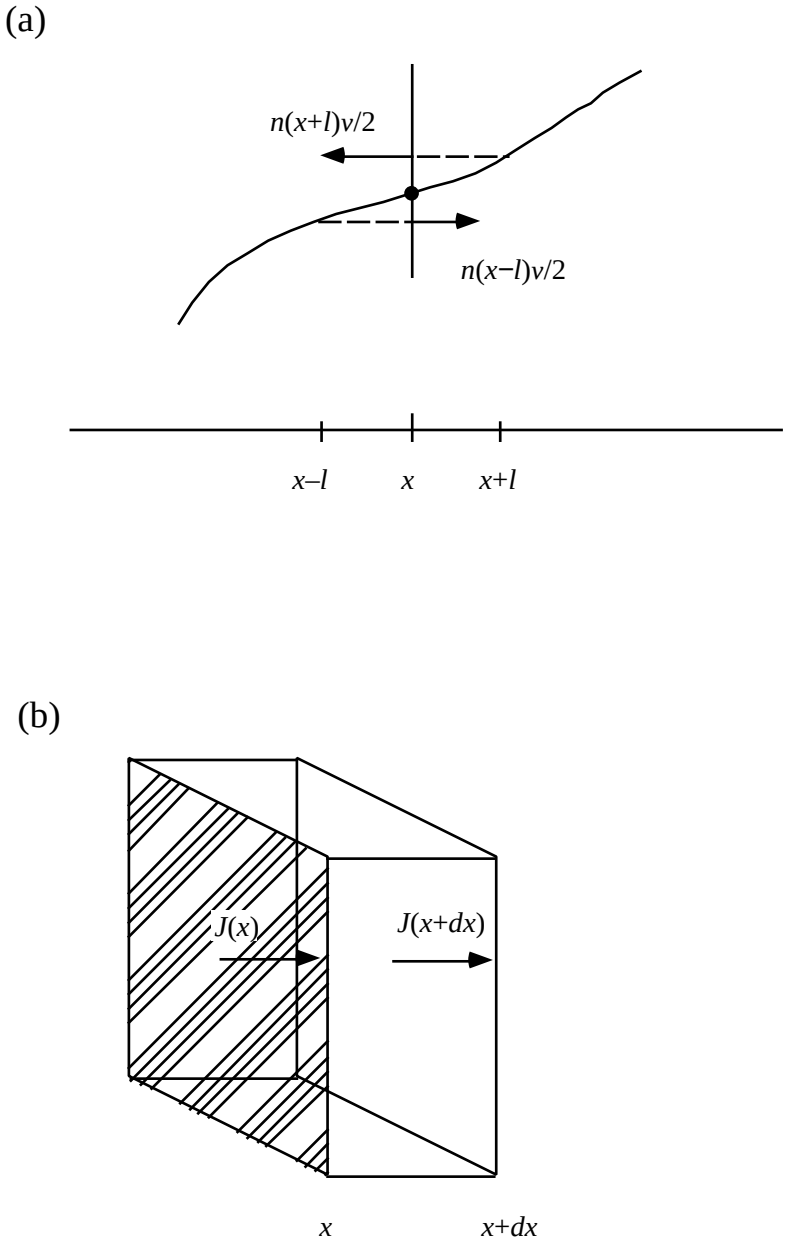


Figure 7.2.12 We derive the diffusion equation using a model consisting of a weakly interacting nonuniform density of particles embedded in a medium. The derivation relates the change in density with time to the spatial variation in density. It takes two steps: (a) the particle current at a point x is related to the spatial variation in density, and (b) the change in density with time is related to the spatial variation in the current. Consult the text for details. ∎

it is controlled by the temperature—and τ is related to the interactions with the fluid or matrix, so neither depend on the density $n(x)$.

When we look at a position x we see molecules traveling to the right and to the left. These molecules originated a distance l to the left and a distance l to the right respectively. At these locations their density was $n(x-l)$ and $n(x+l)$ respectively. Since we expect half of the molecules from $n(x-l)$ to be traveling to the right and half at $n(x+l)$ to be traveling to the left, we infer that the current at x is given by:

$$J(x) = \frac{v}{2}(n(x-l) - n(x+l)) \approx -l\,v\frac{dn(x)}{dx} = -v^2\tau\frac{dn(x)}{dx} \tag{7.2.7}$$

where we have expanded in a Taylor series keeping the first term,and thus assuming that l is small compared to distances over which the density varies significantly.

We want to describe the changes in $n(x)$ as a function of time. To do this we also need the continuity equation that relates the current to the change in density. From Fig. 7.2.12(b) describing the change of density in a small box in terms of the currents at two faces, this is given by

$$\Gamma\Delta x\frac{dn(x;t)}{dt} = \left(-J(x+\Delta x/2;t) + J(x-\Delta x/2;t)\right)\Gamma = -\Gamma\Delta x\frac{dJ(x;t)}{dx} \tag{7.2.8}$$

where Γ is the area of a face and Δx is the length of the side. Combining Eq. (7.2.7) and Eq. (7.2.8) we have the diffusion equation:

$$\frac{dn(x;t)}{dt} = v^2\tau\frac{d^2n(x;t)}{dx^2} \tag{7.2.9}$$

This is generalized to Eq. (7.2.6) when the density varies in three dimensions.

The many assumptions in this derivation can be avoided if we consider Eq. (7.2.6) as an expansion in the density and its derivatives (Question 7.2.4). The right side is the lowest-order term that is not excluded by symmetries of the problem. It controls the longest spatial and temporal behavior. This is the reason for the applicability of the diffusion equation under a large variety of circumstances.

Question 7.2.4 We want to write a differential equation describing the time dependence of the density

$$\frac{dn(x;t)}{dt} = \cdots \tag{7.2.10}$$

in terms of various spatial derivatives—local properties—of the density. Consider including terms that involve up to three derivatives (in three dimensions) of $n(x;t)$:

$$\frac{d}{dx}n(x;t),\quad \frac{d^2}{dx^2}n(x;t),\quad \frac{d^2}{dxdy}n(x;t),\quad \frac{d^3}{dx^3}n(x;t),\quad \frac{d^3}{dx^2dy}n(x;t),\quad \frac{d^3}{dxdydz}n(x;t) \tag{7.2.11}$$

Argue:

a. That of these terms only the second term can be used.

b. There are additional terms involving four derivatives that can be used.

c. That terms of the form

$$n(x;t)\frac{d^2}{dx^2}n(x;t), \quad \left(\frac{d}{dx}n(x;t)\right)^2 \tag{7.2.12}$$

become smaller than the second term in Eq. (7.2.11) when the density is small enough.

d. That terms that do not involve derivatives—a function of $n(x;t)$ itself or a constant—cannot be included if the number of molecules is conserved.

Solution 7.2.4

a. $dn(x;t)/dt$ does not change when we invert any of the spatial coordinates, for example by setting $x \to -x$. Thus any term on the right-hand side of Eq. (7.2.10) must also not change. Since the inversion of x changes the sign of dx no odd derivatives are admissible and only the second term is possible.

b. Additional fourth-order terms are of the form:

$$\frac{d^4}{dx^4}n(x;t), \quad \frac{d^4}{dx^2dy^2}n(x;t) \tag{7.2.13}$$

These terms are corrections to the diffusion equation and must be used if the spatial variations in the density are large enough, or we are concerned about behavior on a small enough length scale.

c. Consider multiplying the density by a factor λ. The terms listed in Eq.(7.2.13) vary as λ^2 while those in Eq.(7.2.11) vary as λ. Thus at low enough density these terms are insignificant.

d. Consider the case of a uniform density $n(x;t) = n_0$. Any function of $n(x;t)$ that does not involve derivatives will give a changing density that must be the same everywhere. A uniform changing density does not conserve the number of molecules. Thus we cannot include such terms. We are implicitly assuming that x itself does not appear in the equation—points in space are indistinguishable before molecules are placed there. Otherwise this argument would not be valid. ∎

Diffusion causes molecules on average to move from higher density regions to lower density regions. This can be readily understood from the random-walk behavior of the molecules and the discussion in Section 1.2. This motion leads to a more uniform density profile. Thus if there is a nonuniform pattern of molecular density initially imposed on a system, diffusion leads to a loss of the pattern through the

smoothing of the density. The key problem in discussing color patterns is identifying how we can cause nonuniform densities to arise out of diffusing molecules. As we remarked before,this is related to the fundamental problem that equilibration generally causes uniformity and lack of structure.

The solution to this problem is through the interaction of more than one type of molecule. Recognizing this was central to the contribution of Turing. The interactions are chemical reactions that change the local densities of molecules. In addition to the reacting molecules, the reactions may involve catalysts that accelerate them. Of particular importance are autocatalyzing reactions where molecules that are reacting are also catalysts. Autocatalysis causes a nonlinear dependence of the reaction rate on the densities. Systems of reacting and diffusing molecules are called reaction-diffusion systems.

7.2.4 *Chemical reactions*

Chemical reactions cause molecular densities to change with time even when there is no diffusion. A reaction may combine different molecules, decompose a molecule into parts or just change the structure of a molecule. We write the general dynamic behavior of the molecular densities using a set of coupled equations of the form:

$$\frac{dn_i(x;t)}{dt} = D\ {}^2 n_i(x;t) + R_i(\{n_j(x;t)\}) \tag{7.2.14}$$

where $R_i(\{n_j(x;t)\})$ is the rate of change in the concentration of a molecular species i due to generation or annihilation in reactions that involve other molecular species. In order to solve such equations,it is necessary to have an expression for $R_i(\{n_j(x;t)\})$ in terms of the densities of the molecules present.

As with diffusion, a discussion of reaction rates requires some simplifying assumptions. In writing Eq.(7.2.14) we have already assumed that the density is not too rapidly varying in space,so that the local reaction rate depends only on the local densities and not their gradients. We will also assume that the diffusion time of a molecule between reactions is large compared to the time of a reaction. This assumption implies that the limiting step in the rate of reaction is the rate at which molecules encounter each other. In order to satisfy this assumption, we need three conditions: that interactions between molecules are short range, that the molecular densities are low and that once the molecules encounter each other the reaction is fast. For simplicity we can think of this as a low density limit. As in the discussion of diffusion, violations of the assumptions can be incorporated in the equations when necessary.

Under these assumptions the rate of a reaction involving molecules A, B and C (with molecular densities n_A, n_B and n_C of the form

$$A + B \rightarrow C \tag{7.2.15}$$

is proportional to the probability of encounter of the reagents—it is proportional to:

$$n_A n_B \tag{7.2.16}$$

This follows from our assumptions because each molecule diffuses and reacts independently of other molecules of the same type. Thus, the probability of a reaction is

proportional to the reactant concentrations. Since one molecule of A and of B disappears for every reaction,and a molecule of C appears, the rate of change of the densities, due to this reaction, are given by:

$$\begin{aligned}\frac{dn_A}{dt} &= -k_1 n_B n_A \\ \frac{dn_B}{dt} &= -k_1 n_B n_A \\ \frac{dn_C}{dt} &= k_1 n_B n_A\end{aligned} \tag{7.2.17}$$

where k_1 is positive, and called a reaction constant.

The reverse reaction

$$A + B \leftarrow C \tag{7.2.18}$$

has a rate which is proportional to n_C. Including this in Eq. (7.3.17) results in the equations:

$$\begin{aligned}\frac{dn_A}{dt} &= -k_1 n_B n_A + k_2 n_C \\ \frac{dn_B}{dt} &= -k_1 n_B n_A + k_2 n_C \\ \frac{dn_C}{dt} &= k_1 n_B n_A - k_2 n_C\end{aligned} \tag{7.2.19}$$

Thus, reactions give rise to differential equations coupling the densities of different molecules.

It is important to emphasize that reactions we write in the form of Eq. (7.2.15) and Eq.(7.2.18) are to be considered elementary reactions that reflect actual molecular encounters. In chemistry, the same notation is often used to describe the net consequence of many reactions. The reaction then reflects only the proportions of molecules involved (stoichiometry). The rate of the reaction is not proportional to reactant density, and therefore must be determined separately.

There are three approximations that can be used to simplify the equations resulting from chemical reactions. These are the condition of quasi-equilibrium,the extreme kinetic regime and the quasi-static regime.

If the two reactions Eq.(7.2.15) and Eq.(7.2.18) are in equilibrium,then the density of A no longer changes with time and we can set Eq. (7.2.19) equal to zero. This gives a relationship between the densities:

$$n_B n_A = k_2' n_C \tag{7.2.20}$$

where $k_2' = k_2/k_1$. When a reaction is close to equilibrium and we disturb the conditions by adding one of them, then the reaction will act to change the densities of the other chemicals to restore equilibrium. If this were the only reaction we were interested in,then the equilibrium would describe all of the dynamics. However, the molecules might be involved in additional reactions that are slower. Then the fast reaction

that restores equilibrium always maintains the chemicals involved in a quasi-equilibrium. Under these conditions we may use the relationship of Eq. (7.2.20) to simplify the system of equations.

The second simplifying circumstance is when the densities of molecules are far from equilibrium. Then one of the two terms in Eq.(7.2.19) will be much larger than the other. In this case we may consider a reaction as proceeding only in one direction. This is the kinetic regime of the reaction, where equilibrium is essentially irrelevant.

The third simplifying circumstance is the quasi-static regime. It is applicable when a quantity is slowly varying on the time scale of observations. The simplest way this can occur is for one of the molecules in a reaction to have a much larger density than the others. Then the change in its density, as compared to the density itself, can be negligible. For example, if the density of C in Eq. (7.2.19) is very large compared to the other molecules,and the value of k_2 is not too large, we may be able to approximate the second term and write, for example:

$$\frac{dn_A}{dt} \approx -k_1 n_B n_A + k_2'' \tag{7.2.21}$$

where $k''_2 = k_2 n_C$ assumes that n_C is approximately constant. This describes a constant source of the molecule A implicitly originating from molecule C. n_C need not be explicitly written when it is essentially constant.

We will be interested in two sets of chemical reactions. The first is the activator-inhibitor system. It represents activation and inhibition more directly, and can be described by

$$\begin{aligned} A &\rightarrow 0 \\ B &\rightarrow 0 \\ 2A + D &\rightarrow 2A + B \\ 2A + C &\rightarrow 3A + C \\ C + B &\leftrightarrow E \end{aligned} \tag{7.2.22}$$

The second is the activator-substrate system. It is simpler and implements the properties of activation and inhibition in a more indirect way to be explained later, and can be described by:

$$\begin{aligned} A &\rightarrow 0 \\ 0 &\rightarrow B \\ 2A + B &\rightarrow 3A \end{aligned} \tag{7.2.23}$$

We discuss simplifications of our treatment of the reactions using the methods discussed above. The discussion will justify the functional form of the differential equations used in the next section.

In both sets of reactions, we have used the convention that 0 represents a chemical species whose density is not of relevance to our discussion. When 0 produces a relevant molecule (e.g., $0 \rightarrow B$) it has a large enough density so that any change is insignificant over the time of observation. This is the quasi-static approximation. We

also denote by 0 a molecule produced by a reaction that is inert (A → 0). This is one way the extreme kinetic limit manifests itself in the reactions. Another way it does so is in all the reactions that have only one direction indicated. The reverse direction is assumed to be irrelevant. There may be other molecules involved in reactions that are not indicated at all. For example we could also write the third reaction in Eq.(7.2.23) as $2A + B + 0 \rightarrow 3A + 0$, where the two 0s indicate molecules whose density is unchanging (the first) or that are inert (the second). Indeed, one of the reactions would definitely not make sense without additional reactants. (Which one?) We will also use the quasi-equilibrium approximation to describe the last reaction in Eq.(7.2.22). Before we describe this, we will discuss the nonlinear reactions that appear in these systems.

Both activator-inhibitor and activator-substrate systems have reaction rates that depend in a nonlinear fashion on molecular densities. The simplest example of a nonlinear dependence is a molecule that reacts with itself:

$$2A \rightarrow B \tag{7.2.24}$$

which would give rise to two coupled equations of the form:

$$\begin{aligned} \frac{dn_A}{dt} &= -k_3 n_A^2 \\ \frac{dn_B}{dt} &= k_4 n_A^2 \end{aligned} \tag{7.2.25}$$

The value of k_3 would be twice as large as k_4 because of the loss of two molecules of A per reaction.

More complex examples of nonlinear dependence result from autocatalysis. First we describe simple catalysis. A catalyst accelerates reactions by creating an additional pathway for the reaction. An example would be:

$$A + B + D \rightarrow C + D \tag{7.2.26}$$

where D is a catalyst since, regardless of intermediate stages, it reappears at the end of the reaction. The density of the catalyst affects the rate of the reaction, but the reaction does not affect the density of the catalyst. An example that appears in Eq.(7.2.22) with A as a catalyst is:

$$2A + D \rightarrow 2A + B \tag{7.2.27}$$

This would give rise to two coupled equations of the form:

$$\begin{aligned} \frac{dn_D}{dt} &= -k_5 n_A^2 n_D \\ \frac{dn_B}{dt} &= k_5 n_A^2 n_D \end{aligned} \tag{7.2.28}$$

Since there is no change in number of A molecules, there is no effect on dn_A/dt.

In autocatalyzed reactions one of the reactants also acts as a catalyst. An example from Eq. (7.2.23) is:

$$2A + B \rightarrow 3A \tag{7.2.29}$$

Each reaction results in the gain of a molecule of *A* and the loss of a molecule of *B*. *A* also acts as a catalyst. The related differential equations take the form:

$$\frac{dn_A}{dt} = k_3 n_A^2 n_B$$
$$\frac{dn_B}{dt} = -k_3 n_A^2 n_B \tag{7.2.30}$$

If two new molecules of A appeared in the reaction, $2A + B \rightarrow 4A$, we would still have the same functional dependence $n_A^2 n_B$. However, the coefficients in the two equations would differ by a factor of 2.

We now consider the last reaction in Eq.(7.2.22) that we will treat using a quasi-equilibrium condition. Some care must be exercised in simplifying equations based upon the interplay between fast processes and the dynamics we are observing. This example is relatively simple because the density of *C* is only affected by the last reaction. It acts as a catalyst in the second to last reaction. We assume that the last reaction is rapid and therefore maintains a relationship between n_C, n_B and n_E similar to that in Eq. (7.2.20):

$$n_B n_C = k_2' n_E \tag{7.2.31}$$

To simplify matters further, we assume that n_E is always very large and effectively constant. Then n_C would be inversely proportional to n_B.

$$n_C = \frac{k_2'''}{n_B} \tag{7.2.32}$$

We can use this relationship in other equations. To illustrate this we write an equation for the time dependence of n_A from the first reaction and the second to last reaction in Eq. (7.2.22):

$$\frac{dn_A}{dt} = -k_1 n_A + k_3 n_A^2 n_C \approx -k_1 n_A + k_3' n_A^2 / n_B \tag{7.2.33}$$

We see that increasing the density of *B* reduces the rate of the reaction of *A* through mediation by *C*. We can say that *B* inhibits the reaction that affects the density of *A*. The next problem is to describe the rate of change of n_B. Since *B* is affected by several reactions in addition to the fast, quasi-equilibrium one, this is more complicated. We can think about the problem as writing a set of equations that no longer contains the variable n_C. While it is not overly difficult to do this, we can simplify matters further by assuming conditions that decouple the behavior of n_B from the quasi-equilibrium equation. This requires that n_B is significantly greater than n_C ($n_C << n_B$). To see how this works we write the complete equations (from Eq. (7.2.22)) that affect n_B and n_C:

$$\frac{dn_B}{dt} = -k_4 n_B n_C + k_5 n_E + (k_3 n_A^2 n_D - k_2 n_B) \tag{7.2.34}$$

$$\frac{dn_C}{dt} = -k_4 n_B n_C + k_5 n_E \tag{7.2.35}$$

The density n_C changes only through the reaction that is in quasi-equilibrium. n_B has the same terms, but also the terms in parenthesis reflecting the additional reactions that include *B*. If $n_C << n_B$ then any change in n_C is also much smaller than n_B. Thus we can neglect the first two terms in the rate of change of n_B which are the same as the rate of change of n_C. Then we are left with only the terms in parenthesis:

$$\frac{dn_B}{dt} = k_3 n_A^2 n_D - k_2 n_B \tag{7.2.36}$$

A more complete treatment is discussed in Questions 7.2.5 and 7.2.6.

Question 7.2.5 Write an expression instead of Eq. (7.2.32) for the dependence of n_C on n_B when we cannot assume that n_E is unchanging.

Solution 7.2.5 In order to obtain the more general form of Eq.(7.2.32) we must recognize that the sum $n_E + n_C$ is conserved in the reactions of Eq. (7.2.22). We can define this sum to be n_0 and write the quasi-equilibrium condition Eq. (7.2.31) as:

$$n_B n_C = k_2'(n_0 - n_C) \tag{7.2.37}$$

or

$$n_C = k_2' n_0 / (n_B + k_2') \tag{7.2.38}$$

We see that as long as n_B is larger than k_2' this correction can be ignored. The correction will be important when we do simulations later because it is unphysical that the rate of change of n_A given in Eq.(7.2.33) diverges when the density of n_B is small. ∎

Question 7.2.6 Derive an equation instead of Eq.(7.2.36) that incorporates an approximate quasi-equilibrium relationship but doesn't assume $n_C << n_B$. Assume n_E is essentially constant.

Solution 7.2.6 We start from the quasi-equilibrium relationship Eq.(7.2.31). To use this relationship we recognize that the equality is not exact, but holds approximately. The difference between the two sides, which appears in Eq. (7.2.35), ensures that n_C changes when n_B does. Any change in n_B must be matched by a change in n_C to maintain the quasi-equilibrium relationship itself. Thus an incremental change of Eq. (7.2.31) can be written:

$$n_B dn_C + n_C dn_B \approx 0 \tag{7.2.39}$$

where we have assumed n_E is essentially constant. Dividing by a time increment *dt* and using the approximate quasi-equilibrium relationship we relate the rate of change of n_C to that of n_B:

$$\frac{dn_C}{dt} \approx -\frac{k_2'''}{n_B^2}\frac{dn_B}{dt} \tag{7.2.40}$$

We use this expression instead of the first two terms on the right side of Eq. (7.2.34):

$$\frac{dn_B}{dt} \approx -\frac{k_2'''}{n_B^2}\frac{dn_B}{dt} + (k_3 n_A^2 n_D - k_2 n_B) \tag{7.2.41}$$

or

$$\left(1 + \frac{k_2''}{n_B^2}\right)\frac{dn_B}{dt} \approx (k_3 n_A^2 n_D - k_2 n_B) \tag{7.2.42}$$

or finally:

$$\frac{dn_B}{dt} \approx \frac{1}{1 + k_2''' n_B^{-2}}(k_3 n_A^2 n_D - k_2 n_B) \tag{7.2.43}$$

The precise conditions under which this equation is valid can be understood by recognizing that Eq.(7.2.40) can be obtained from the time derivative of Eq. (7.2.35) by neglecting the second time derivative of n_C. ∎

7.2.5 *Pattern formation in reaction-diffusion systems*

The combination of reaction and diffusion terms in a differential equation can give rise to pattern formation under particular circumstances. Ultimately the source of the pattern formation may be the same as that of the CA rules—short-range activation and long-range inhibition. However, this is not as transparent in the differential equation form. There are two ways to think about the formation of differential equation patterns. The first is a conceptual one that connects with activation and inhibition. The second is through the analytic properties of the differential equations that can give rise to a pattern. To understand the conceptual approach, we must relate the notion of action at a distance of the CA model to the reaction-diffusion model. The influence of one molecular species over a distance is achieved by diffusion. Typically, when diffusion is faster the influence is longer range. Since there are two processes—activation and inhibition—that occur over different ranges, it makes sense to consider the effects of two types of molecules, one with a short-range influence corresponding to a small diffusion constant, and one with a long-range influence corresponding to a large diffusion constant. Activation is a process by which one cell produces a signal molecule that causes other cells around it to produce the same molecule. From the point of view of the molecules, this is a self-catalyzing reaction that causes a nonlinear increase in the molecule density. Thus we expect that the molecule with a small diffusion constant autocatalyzes a reaction that increases its own density. Cell pigment is then coupled to its density. The second molecule, with a longer-range influence, must perform an inhibi-

tion of the reaction that forms the first molecule. We will see that the equations developed to demonstrate pattern formation have these properties.

Efforts have been made to construct models of actual physiological reaction-diffusion processes. It is to be expected that such systems involve more than two types of molecules,though quasi-static,kinetic and quasi-equilibrium approximations may allow their description to be simplified. We will discuss two sets of equations that are not specifically obtained from the physiology of pattern formation but are used to illustrate how the patterns can form. The equations have only two types of molecules A and B whose density $n_A(x,y;t)$ and $n_B(x,y;t)$ in space and time we write for simplicity as $a(\mathbf{x};t)$ and $b(\mathbf{x};t)$. The molecules diffuse with different diffusion constants D_a and D_b. The differential equations describing their behavior can be written generally as

$$\begin{aligned}\frac{da(\mathbf{x};t)}{dt} &= D_a \ ^2a(\mathbf{x};t) + f(a(\mathbf{x};t),b(\mathbf{x};t))\\ \frac{db(\mathbf{x};t)}{dt} &= D_b \ ^2b(\mathbf{x};t) + g(a(\mathbf{x};t),b(\mathbf{x};t))\end{aligned} \tag{7.2.44}$$

The functions $f(a,b)$ and $g(a,b)$ reflect the effects of chemical reactions. They describe the time dependence of the densities when the density is uniform.

We now write down and simulate two sets of equations that form patterns. The first set of equations may be obtained from the activator-inhibitor reactions in Eq. (7.2.22) as discussed in the previous section (see Question 7.2.7). They are described by:

$$\begin{aligned}f(a,b) &= k_1a^2/b - k_2a\\ g(a,b) &= k_3a^2 - k_4b\end{aligned} \tag{7.2.45}$$

The first term k_1a^2/b describes the autocatalytic formation of the activator A which is inhibited by the presence of B. The inhibitor B is produced by A in the term k_3a^2. If the molecules of B are rapidly diffusing, the creation of B in regions where a is large causes long-range inhibition of the formation of A. The densities of both A and B are limited by decay processes responsible for the second term in each equation. Patterns formed from this set of equations are shown in Figs.7.2.13 and 7.2.14. We will discuss the methodology of these simulations in greater detail below.

The second set of equations that we use to generate patterns may be obtained from the activator-substrate reactions in Eq. (7.2.23) (see Question 7.2.8):

$$\begin{aligned}f(a,b) &= k_1a^2b - k_2a\\ g(a,b) &= k_3 - k_4a^2b\end{aligned} \tag{7.2.46}$$

In this set of reactions,the presence of B is necessary for the autocatalytic reaction that creates A, as is evident in the expression k_1a^2b. The same reaction causes the disappearance of B and the formation of A. B is spontaneously created by a process, given by k_3, that is independent of the density of A or B. Finally, the density of A is limited by decay, as evident in the term $-k_2a$. We can consider the autocatalytic increase of A as local self-activation. Long-range inhibition arises when the diffusion constant of B

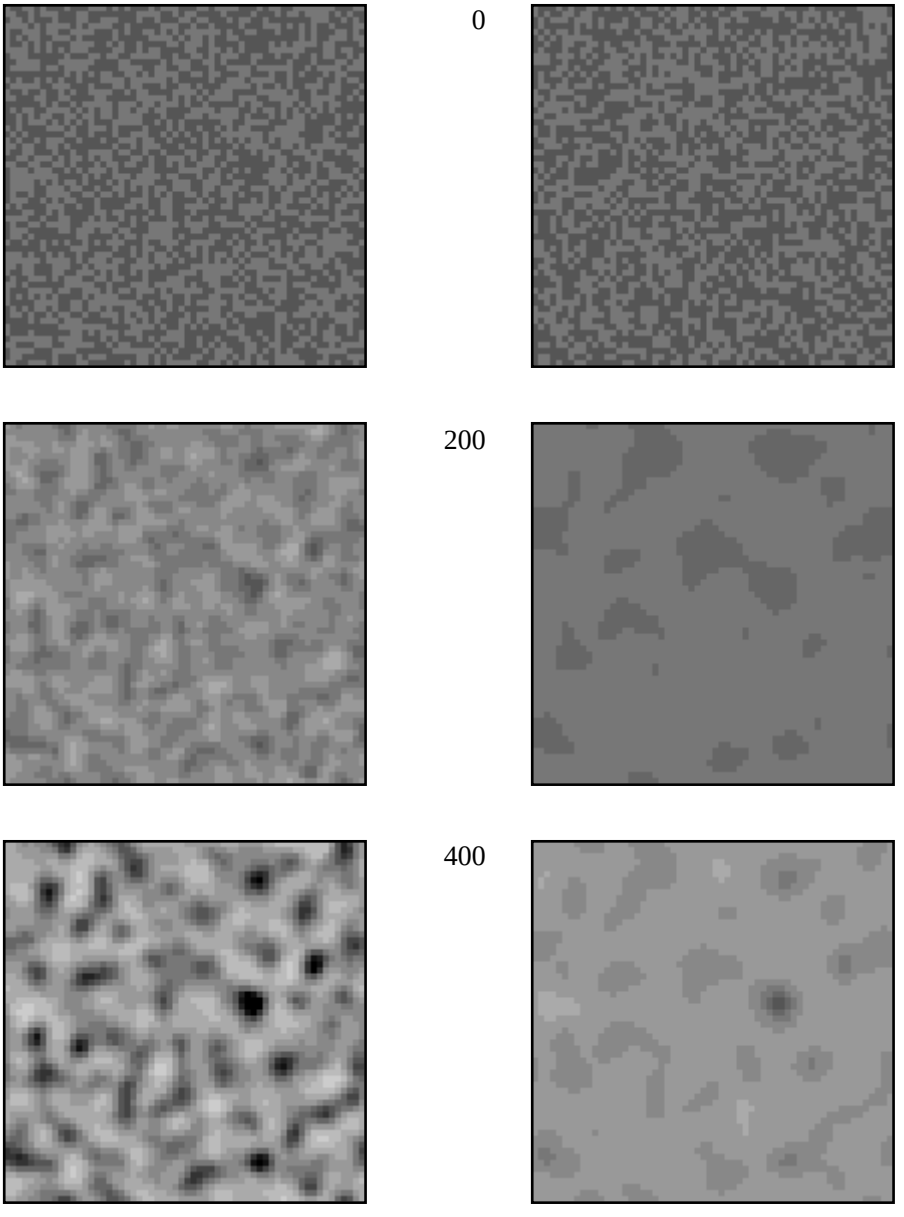
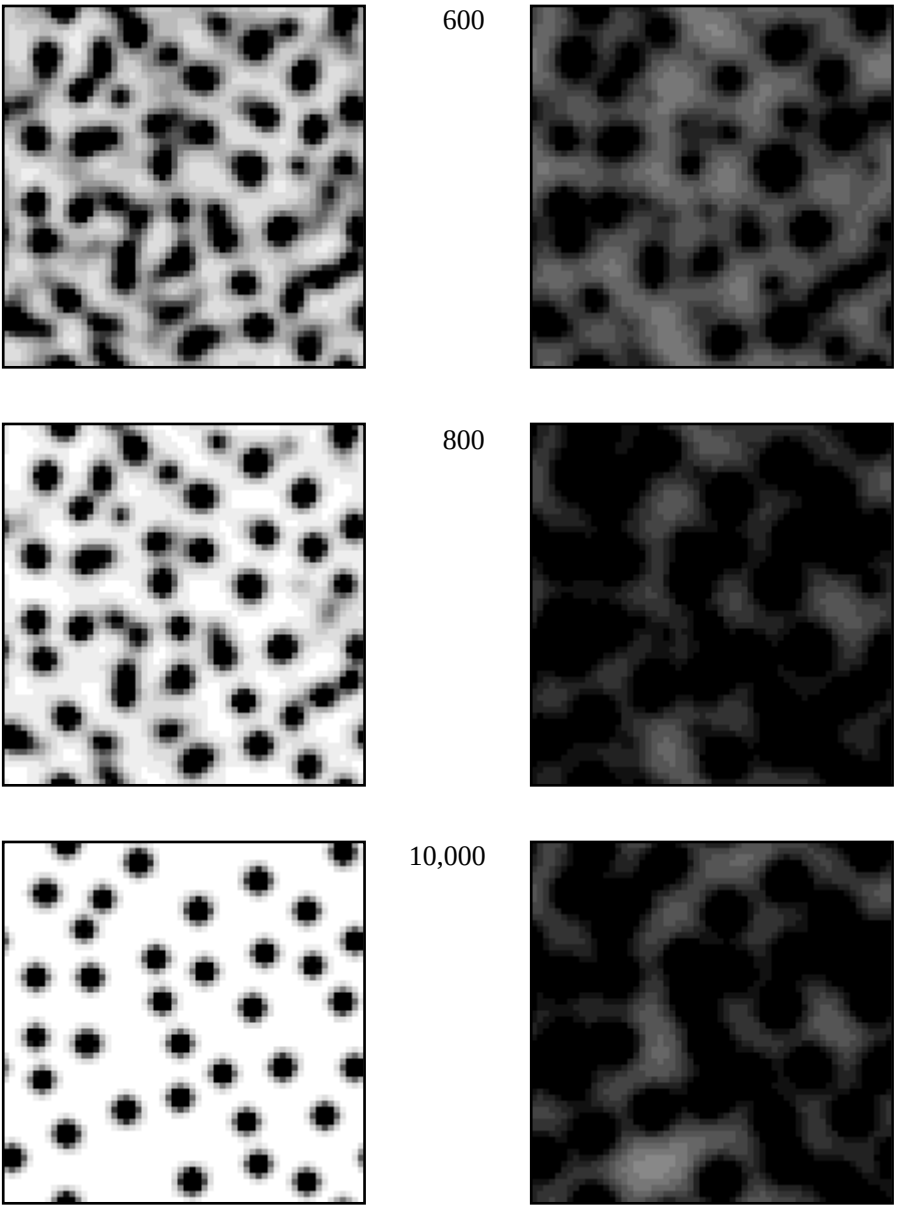


Figure 7.2.13 Simulations of the first set of reaction-diffusion equations, the activator-inhibitor system. At each time two panels are shown. The left panel shows the density *a* of the activator *A*. The right panel shows the density *b* of the inhibitor *B*. The parameters were chosen as described in the text with $k_1 = k_2 = k_3 = k_4 = 1$. The initial conditions, shown as the first panel, consist of density values either of 1 or of 1.3 placed randomly with equal probability. The same initial conditions are used for Figs. 7.2.14 - 7.2.16. Note that since *B* is created by *A* they both have maxima and minima at the same locations. The more rapid diffu-



sion of B causes the regions around the maxima to be depleted of A. The plots show the density using a representation (gray scale) that uses gray values ranging from white for 0 to black for 2. The figures are labeled by the time in units of updates. Since our convention is that the time per update is $\Delta t = 0.01$ the frame marked 200 would correspond to $t = 2$. A steady state is essentially reached by 10,000 updates. This was verified and used throughout for the other reaction-diffusion simulations in Figs. 7.2.14–7.2.16. Note the difference between this and the number of updates (20) necessary for the CA models of Section 7.2.2. ∎

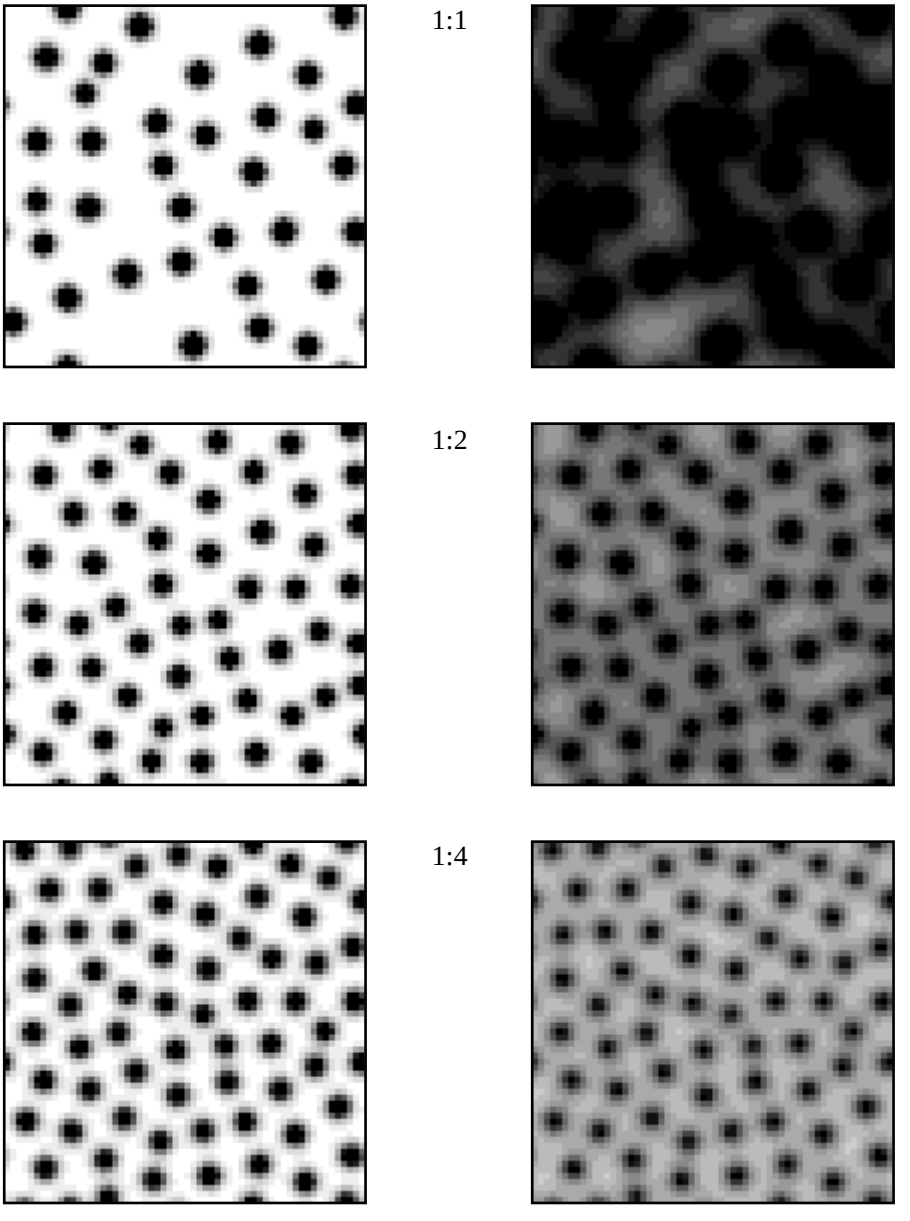


Figure 7.2.14 Simulations of the activator-inhibitor reaction-diffusion system for different values of the reaction constants. The left panels show the density a of the activator A. The right panels show the density b of the inhibitor B. All frames show the steady-state result after 10,000 updates. The parameters for the frames are $k_1 = k_2 = 1$ and $k_3 = k_4 = 1, 2, 4$ respectively. The parameters for the top frames are the same as Fig. 7.2.13 and reproduce the last time step of that figure. ∎

is much larger than the diffusion constant of A. Because B moves rapidly and is consumed by reaction with A, the density of B is depleted not only where A is high in density, but also in the surrounding region. Since B is necessary for the creation of A, this inhibits the formation of A in this larger region. Patterns formed from this set of equations are shown in Fig. 7.2.15. In the activator-inhibitor set of reactions,the maxima of b occur at the same locations as the maxima of a. In the activator-substrate system, the minima of b are at the maxima of a.

Question 7.2.7 Identify the relevant equations and approximations from Section 7.2.4 used to obtain Eq. (7.2.45) from the reactions in Eq. (7.2.22).

Solution 7.2.7 The two most directly relevant equations are Eq. (7.2.33) and Eq. (7.2.36). The approximations leading to them are relevant, including the quasi-equilibrium approximation for the last reaction in Eq.(7.2.22). The only additional modification is that n_D, which plays no real role in the discussion of the last section, is assumed to be essentially unchanging. ∎

Question 7.2.8 Identify the approximations used to obtain Eq.(7.2.46) from the reactions in Eq.(7.2.23). There is an inconsistency between the reactions and the equations. In the activator-substrate system the reaction

$$2A + B \rightarrow 3A \tag{7.2.47}$$

is the only reaction that is responsible for two terms in the differential equations. These two terms in Eq. (7.2.46) have the coefficients k_1 and k_4. This would mean that $k_1 = k_4$, since one A is gained and one B is lost. Describe a modified reaction in which $k_1 = 2k_4$ (easy) and a reaction in which $k_4 = 2k_1$(hard). One of our assumed cases in the simulations corresponds to the latter.

Solution 7.2.8 For the case $k_1 = 2k_4$ we produce twice as many A in each reaction as B is lost, this can be done using:

$$2A + B \rightarrow 4A + C \tag{7.2.48}$$

The difficulty in the case $k_4 = 2k_1$ is that the left side of the equation must have only one B, but we want to make twice as many B disappear as A appears. To do this we need to have A be a composite molecule formed by a fast reaction from two equivalent parts (ligands) that bind together to form a complete A. We call each part D, then we have the reactions:

$$\begin{aligned} 2A + B &\rightarrow 2A + D + C \\ 2D &\rightarrow A \end{aligned} \tag{7.2.49}$$

where the second is a fast reaction. This combination gives the desired result.

Another possible solution is to use two catalyzed reactions. One causes A to appear and is catalyzed by B. The second causes B to disappear and is catalyzed by $2A$. Then the coefficients can be set independently. This suggests some of the subtlety necessary to create actual pattern-forming systems. ∎

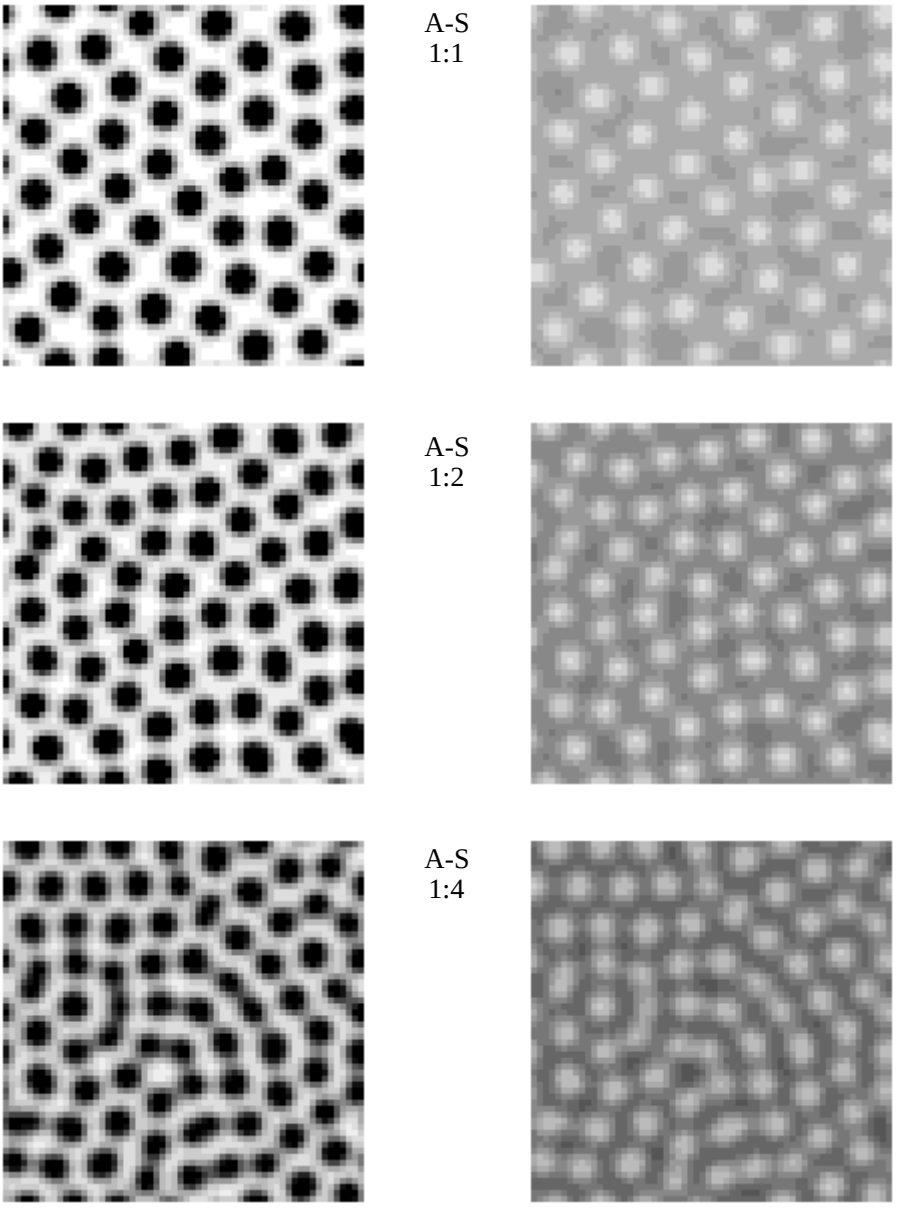


Figure 7.2.15 Similar to Fig. 7.2.14, but for the activator-substrate system. The right panels show the substrate *B*, which is consumed by the activator *A*. *B* is depleted and has its minima where *A* has its maxima. Due to the more rapid diffusion of *B*, it is depleted in a region around maxima of *A*. Thus, the growth of *A* is inhibited in regions surrounded by maxima of *A*. ∎

Eqs. (7.2.45) and (7.2.46) involve many parameters,six different ones including the two diffusion constants and four reaction constants. Exploring the six-dimensional parameter space would involve much effort. Exploring large dimensional spaces to discover particular pattern-forming regions of the space can give insight into the difficulty of evolutionary processes that form these systems. However, we can significantly simplify our problem mathematically by recognizing that each of the densities a and b and the variables x and t can be measured in convenient units. By normalizing these variables we do not change the form of the pattern, only its scale. Full use of this would reduce the number of independent parameters to only two. For our simulations,the time and length scale must be related to the time step and lattice size. However, we can conveniently scale the densities a and b.

It is easier to scale the densities if we make use of the observation that these equations have a solution that is uniform and does not change in time. This solution, obtained by setting $f(a,b) = g(a,b) = 0$, is unstable in the parameter domain in which patterns form. Adding any small perturbation leads to the formation of a pattern. We will discuss this in more detail at the end of this section. In the meantime we use the uniform solution to choose coefficients—patterns typically consist of positive and negative excursions from the unstable uniform solution. By normalizing the coefficients, we can set the uniform solution so that it is $a = b = 1$. For both sets of equations, this imposes the same relationships between the coefficients:

$$\begin{aligned} k_1 &= k_2 \\ k_3 &= k_4 \end{aligned} \tag{7.2.50}$$

Using these relationships also makes it easier to display simulations,since we can use a consistent scale for all plots of the densities. All of the figures showing density plots of the patterns are formed using a scale that begins with white at 0 and ends with black at 2.

In simulating the behavior of these differential equations, we can use a finite difference representation of the diffusion operator:

$$\frac{d^2a(x)}{dx^2} \rightarrow \frac{1}{\Delta x^2}\big(a(i+1)+a(i-1)-2a(i)\big) \tag{7.2.51}$$

or in two dimensions:

$$\frac{d^2a(\mathbf{x})}{dx^2}+\frac{d^2a(\mathbf{x})}{dy^2} \rightarrow \frac{1}{\Delta x^2}\big(a(i+1,j)+a(i-1,j)+a(i,j+1)+a(i,j-1)-4a(i,j)\big) \tag{7.2.52}$$

The time derivative is represented as a time difference:

$$\frac{da(t)}{dt} \rightarrow \frac{1}{\Delta t}\big(a(t)-a(t-1)\big) \tag{7.2.53}$$

where we use t also as the discrete time index. These substitutions return us to a CA consistent with a random-walk model of molecular motion. It has the form:

$$
\begin{aligned}
a(i,j;t+1) &= a(i,j;t) + \Delta t\, f(a(i,j;t),b(i,j;t)) \\
&\quad + \frac{\Delta t}{\Delta x^2} D_a \left(a(i+1,j;t) + a(i-1,j;t) + a(i,j+1;t) + a(i,j-1;t) - 4a(i,j;t)\right) \\
b(i,j;t+1) &= b(i,j;t) + \Delta t\, g(a(i,j;t),b(i,j;t)) \\
&\quad + \frac{\Delta t}{\Delta x^2} D_b \left(b(i+1,j;t) + b(i-1,j;t) + b(i,j+1;t) + b(i,j-1;t) - 4b(i,j;t)\right)
\end{aligned}
\tag{7.2.54}
$$

The choice of Δt and Δx is coupled to the choice of the remaining coefficients—the reaction constants k_1 and k_3 and the value of the two diffusion constants D_a and D_b. Their value determines the characteristic time to equilibration and the length scale of the pattern that is found. The time scale must be set so that significant changes do not happen in a single increment, because otherwise the differential equation is not being correctly approximated, and oscillatory or chaotic dynamics of iterative maps may occur. The inherent time scale of the system is set by the amount of time it takes for a typical molecular density to change significantly. If we assume that our reaction constants k_1 and k_3 are set approximately equal to one, and we have already chosen the characteristic density of both reactants to be one,then the time for the characteristic density to change is also one. We must ensure that this is much larger than the time interval Δt so we choose $\Delta t = 0.01$. For the simulations we choose the length scale to be approximately one lattice constant, so we set $\Delta x = 1$.

How should we choose the values of the diffusion constants D_a and D_b ? We can set their relative values by noting that the range of diffusion $\sqrt{Dt}$ is proportional to the square root of the diffusion constant. In the CA models, we used a ratio of activation range to inhibition range of 6:1. Thus we would like the diffusion constants $D_b : D_a$ to be approximately 36:1 with D_a approximately 1. For the simulations, a ratio of 40:1 was used with $D_a = 0.5$ and $D_b = 20$. $D_b = 20$ was used instead of $D_b = 40$, because for this value the coefficient of $b(i,j;t)$ in Eq. (7.2.54) is greater than 1 (it is −1.6) which causes numerical instabilities (see also Question 7.2.9).

With most of the parameters determined,the only remaining choice is the relative values of the reaction constants k_1 and k_3 with both approximately one. We fix $k_1 = 1$,and vary k_3. Not all values of k_3 produce patterns. In Fig. 7.2.13 patterns formed from the activator-inhibitor system are shown for $k_3 = 1,2,4$. For smaller values of k_3, the spots become sparser as is evident already from the behavior at $k_3 = 1$. For higher values of k_3 ($k_3 = 8$),the pattern disappears and a uniform solution of the differential equation becomes the steady-state result. Simulations for the same values of $k_3 = 1,2,4$ are shown in Fig. 7.2.14 for the activator-substrate system. However, in this case we see that at low values of k_3 the spots become slightly bigger but not significantly sparser. For still lower values of k_3, the simulations, as described above, become unstable and do not arrive at a steady-state result. For higher values of k_3 ($k_3 = 8$),a uniform solution becomes stable. An analytic approach to understanding the pattern-forming range of k_3, and incidentally why both sets of equations have a similar pattern-forming behavior, is described in Questions 7.2.7 and 7.2.8.

The finite difference form of the differential equations in Eq. (7.2.54) is a CA. This CA is both simpler and more elaborate than the CA in Section 7.2.2. Here the interactions between cells are nearest neighbor and the variables at every site are two real numbers—a major part of the pattern-forming behavior arises from the on-site part of the rule. In Section 7.2.2 the interactions were longer range and each cell had only a single binary variable—the pattern formation arose from the interactions. We note that CA rules that are derived from differential equations are designed to be studied in the limit where the cell size is small enough that granularity does not affect the result. This is not necessarily the case with all CA rules;however, in the case of pattern formation,a similar limit should be taken where the cell size is small compared to the typical size of the pattern.

Question 7.2.9 The parameters of the differential equations that give rise to patterns must,in biological systems,arise out of the properties of the molecules involved. If we assume that simple diffusion applies, the diffusion constant arises largely from the volume of the molecule, so the slow diffuser must be much larger than the fast diffuser. Discuss the practicality of the activator-inhibitor or activator-substrate systems simulated here.

Solution 7.2.9 Using Stokes' law (see also Section 5.2) for spherical molecules,the diffusion constant is inversely proportional to the cube root of the volume. For simplicity we can assume the volume is approximately proportional to the mass. Since the diffusion constants were set to have a ratio of 40:1,the masses must have a ratio of 64,000 or approximately 10^5. Recall that the characteristic distance traveled is proportional to the square root of the diffusion constant, which is inversely proportional to the cube root of the mass. Thus the characteristic distance traveled is inversely proportional to the sixth root of the mass—a very weak function of the mass.

Since the fast diffuser must be complicated enough to participate in well-defined ways in specific reactions, we can not expect it to be easy to design a small molecule to do this. If the small molecule is itself large,the large molecule must be huge. Thus, either the slow diffuser must be a behemoth or some other approach must be taken.One solution is that the slow diffuser is actually a cell rather than a molecule (see Section 7.2.7). Another possibility is that other effects, such as reactions that temporarily bind molecules, reduce its diffusion rate. ∎

Using the reaction-diffusion equations and the chosen parameters, the patterns formed are those of spots. We have seen from the discussion of the CA activation-inhibition models in Section 7.2.2 that there are several ways to cause such patterns to form stripes.One way (Question 7.2.2) is to change the relative strength of the inhibition compared to the activation. In the reaction-diffusion systems,the same terms in the differential equations are responsible for both activation and inhibition (k_1 in Eq. (7.2.44) and k_1, k_4 together in Eq. (7.2.45)). Thus it does not appear possible to

control separately the activation and inhibition. However, these terms describe activation when a is large and inhibition when a is small. Thus we can vary their relative strength by introducing an additional density dependence to these terms that reduces the activation at high values of a and maintains the inhibition at low values of a. For the first set of equations (activator-inhibitor):

$$\begin{aligned} f(a,b) &= k_1 a^2 / b(1 + k_5 a^2) - k_2 a \\ g(a,b) &= k_3 a^2 - k_4 b \end{aligned} \tag{7.2.55}$$

For the second set (activator-substrate):

$$\begin{aligned} f(a,b) &= k_1 a^2 b / (1 + k_5 a^2) - k_2 a \\ g(a,b) &= k_3 - k_4 a^2 b / (1 + k_5 a^2) \end{aligned} \tag{7.2.56}$$

While we do not discuss the possible chemical origins of this modification in detail, we can understand it as a saturation (effectively an inhibition) of the autocatalytic reaction in the presence of high densities of the activator. It could be caused by an additional inhibitor whose density is tied to a. Patterns formed from these equations in Fig. 7.2.16 show the formation of stripes.

In summary, we see that the conditions under which patterns can be generated include cases where there are two types of molecules, one diffusing rapidly and the other slowly. The slow diffuser A autocatalyzes a reaction that increases its own density. The fast diffuser B reacts with the slow diffuser and decreases the density of A in the vicinity of a high-density region of A. This results in patterns like that of the activation-inhibition CA model in the previous section. The primary difference between the two sets of differential equations is that the fast diffuser B acts to inhibit in two distinct ways, in the activator-inhibitor system through its presence, and in the activator-substrate system through its absence (depletion).

The discussion of these equations in terms of activation and inhibition can be augmented by a discussion of their analytic properties. Diffusion in the absence of reactions causes the density to become uniform and patterns are not possible. What are the mathematical conditions under which patterns will form when there are reactions? Central to our understanding of the formation of patterns is the recognition that a uniform solution of the equations continues to exist even when patterns are formed. However, this uniform solution is unstable. This means that adding a small nonuniform density (perturbation) to the uniform solution will cause the system to evolve to a pattern such as those shown in the figures. An analytic study of the stability of the uniform solution is known as linear stability analysis. Using a linear expansion of the equations around the uniform solution, we can determine if it is stable. When it is not stable then the quadratic terms become important in determining the solution, which may be a nonuniform pattern.

We can take the analysis one step further by recalling that a key aspect of the pattern is the existence of a length scale characteristic of the distance between spots. This length scale arises even though a differential equation (unlike the CA) has no cellular length scale;it is also independent of the size of the system. The characteristic length

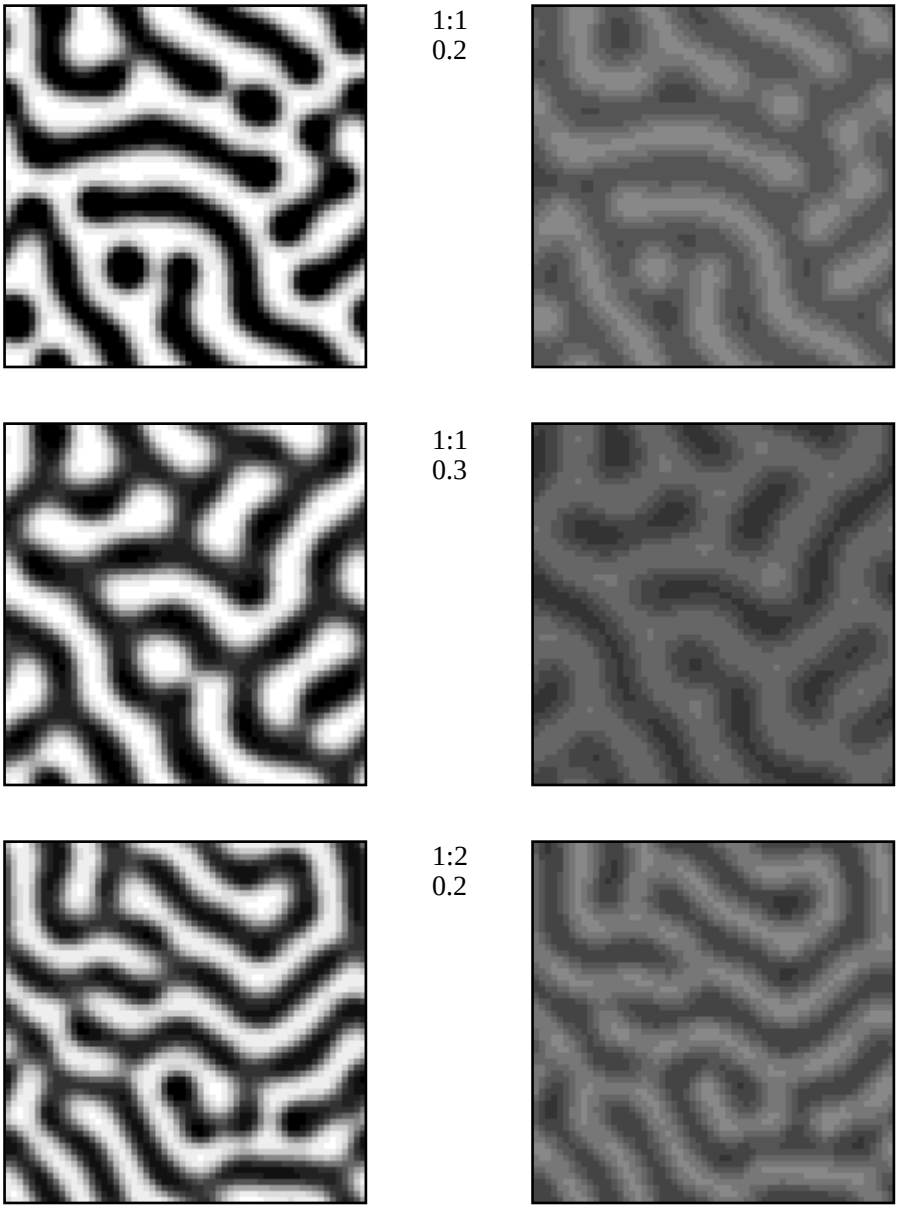


Figure7.2.16 The addition of a parameter that causes the rate of growth of A to be decreased at high density of A and increased at low density causes the formation of stripes in both the activator-inhibitor (this page) and activator-substrate (p. 670) systems. The parameter values shown are: for all cases $k_1 = k_2 = 1$, for top $k_3 = k_4 = 1$, $k_5 = 0.2$, for middle $k_3 = k_4 = 1$, $k_5 = 0.3$ and for bottom $k_3 = k_4 = 2$, $k_5 = 0.2$. ∎

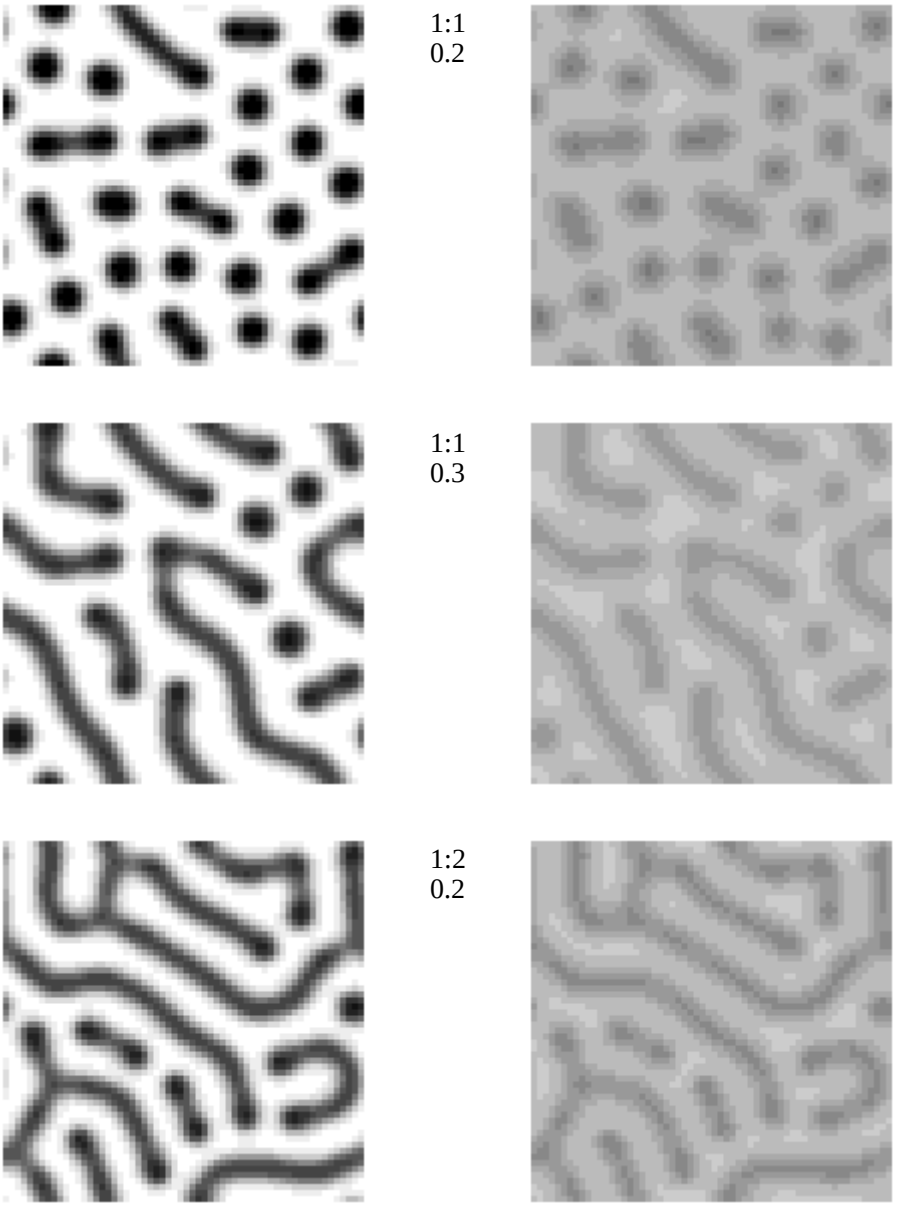


Figure 7.2.16 (*continued*)

scale arises because of the nature of the instability of the uniform solution. Instead of being unstable to all perturbations, the system is only unstable to perturbations of a range of length scales that characterize the patterns. Using the linear expansion around the stable solution, we can identify the range of length scales over which it is unstable to perturbations,and thus identify whether a pattern will form (or for what range of parameters a pattern may be expected to form), and what its characteristic length scale should be. This analysis is discussed in Questions 7.2.10 and 7.2.11.

Question 7.2.10 Patterns are generated when a differential equation has a uniform steady-state solution that is unstable to perturbations at the length scale of the pattern. The instability means that a small addition to the uniform solution grows in time until it is stopped by a process that limits the continued growth. To perform an analytical investigation of the reaction-diffusion equations, expand the reaction part of the reaction-diffusion equations $f(a,b)$, $g(a,b)$ around the uniform solutions in the form

$$\begin{aligned} a &= 1 + u \\ b &= 1 \mp v \end{aligned} \tag{7.2.57}$$

using the –(+) sign for the activator-inhibitor (activator-substrate) system of equations. Then write differential equations for the time evolution of u and v. It is only necessary to keep the linear terms.

Solution 7.2.10 We expand $f(a,b)$, $g(a,b)$ to second order. We use only the first-order terms, but the second-order terms will illustrate a point. Inserting Eq. (7.2.57) and expanding the activator-inhibitor set of equations gives:

$$\begin{aligned} f(1+u,1-v) &= k_1(1+u)^2/(1-v) - k_1(1+u) + \\ &= -k_1(1+u) + k_1(1+2u+u^2)(1+v+v^2+\ldots) \\ &= -k_1(1+u) + k_1(1+2u+u^2+v+2uv+v^2+\ldots) \\ &= k_1(u+v) + k_1(u+v)^2 + \ldots \\ g(1+u,1-v) &= k_3(1+u)^2 - k_3(1-v) \\ &= k_3(2u+v) + k_3u^2 \end{aligned} \tag{7.2.58}$$

For the activator-substrate set of equations, we have:

$$\begin{aligned} f(1+u,1+v) &= k_1(1+u)^2(1+v) - k_1(1+u) \\ &= -k_1(1+u) + k_1(1+2u+u^2+v+2uv+u^2v) \\ &= k_1(u+v) + k_1(u^2+2uv) + \ldots \\ g(1+u,1+v) &= k_3 - k_3(1+u)^2(1+v) \\ &= -k_3(2u+v) - k_3(u^2+2uv) + \ldots \end{aligned} \tag{7.2.59}$$

The differential equations for u and v are obtained by inserting Eq.(7.2.57) and Eq.(7.2.58) or Eq.(7.2.59) into Eq.(7.2.44). After substitution we switch signs in the second equation, when necessary, to obtain:

$$\frac{du(\mathbf{x};t)}{dt} = D_a \nabla^2 u(\mathbf{x};t) + f(1+u(\mathbf{x};t), 1 \mp v(\mathbf{x};t)) \approx D_a \nabla^2 u(\mathbf{x};t) + k_1(u(\mathbf{x};t) + v(\mathbf{x};t))$$
$$\frac{dv(\mathbf{x};t)}{dt} = D_b \nabla^2 v(\mathbf{x};t) \mp g(1+u(\mathbf{x};t), 1 \mp v(\mathbf{x};t)) \approx D_b \nabla^2 v(\mathbf{x};t) - k_3(2u(\mathbf{x};t) + v(\mathbf{x};t)) \tag{7.2.60}$$

We have written only the first-order terms, which are the same in both sets of equations. The second-order terms are not the same. The equivalence of the first-order terms in part explains the similarity in the results obtained by simulating the two sets of equations. The inequivalence of the second-order terms is responsible in large part for the differences. ∎

Question 7.2.11 Eq. (7.2.60) consists of two coupled linear differential equations. For a uniform solution where u and v are independent of x, the solution must either be a growing exponential or a decaying exponential. The two possibilities correspond to an unstable or stable uniform solution of the original equations. We can also consider nonuniform solutions by using the trial solutions:

$$\begin{aligned} u(\mathbf{x};t) &= u_0 e^{\lambda t} \sin(\kappa x + \phi) \\ v(\mathbf{x};t) &= v_0 e^{\lambda t} \sin(\kappa x + \phi) \end{aligned} \tag{7.2.61}$$

A one-dimensional spatial variation has been assumed,since there is no y dependence. Substitute and find possible values of λ. Plot the real part of λ as a function of κ for the parameter values used in the simulations above. When the real part of λ is positive,the uniform solution is unstable; when the real part is negative, the uniform solution is stable.

Solution 7.2.11 Substituting Eq. (7.2.61) into Eq. (7.2.60) we have:

$$\begin{aligned} \lambda u_0 &= -\kappa^2 D_a u_0 + k_1(u_0 + v_0) = (-D_a\kappa^2 + k_1)u_0 + k_1 v_0 \\ \lambda v_0 &= -\kappa^2 D_b v_0 - k_3(2u_0 + v_0) = -2k_3 u_0 + (-D_b\kappa^2 - k_3)v_0 \end{aligned} \tag{7.2.62}$$

To determine the solutions, we must find eigenvectors and eigenvalues of the matrix:

$$\begin{pmatrix} -D_a\kappa^2 + k_1 & k_1 \\ -2k_3 & -D_b\kappa^2 - k_3 \end{pmatrix} \tag{7.2.63}$$

The eigenvalues, which are the possible values of λ, can be obtained with some algebra:

$$\lambda_\pm = \frac{1}{2}\left(-\kappa^2(D_a + D_b) + k_1 - k_3 \pm \sqrt{(-\kappa^2(D_a - D_b) + k_1 + k_3)^2 - 8k_1k_3}\right) \tag{7.2.64}$$

We could find the solutions (values of u_0 and v_0).Our objective,however, is only to consider the eigenvalues $\lambda_{\pm}$. Their real part determines whether the solutions grow or decay. If they decay, then the uniform solution of the original equations $a = b = 1$ is stable and no pattern will form. If one of the solutions grows,then the system will form a pattern. Without analyzing these eigenvalues in great detail, we can plot their values for the parameters used in the simulations to form patterns as a function of $1/\kappa$, which is proportional to the length scale of the perturbation. This is done in Fig. 7.2.17. We see that the real part of λ_+ is positive for a range of values around unity but is negative both at $1/\kappa = 0$ and $1/\kappa = \infty$. This means there is a limited range of length scales at which the equations are unstable, and this range determines the size of the pattern that is formed. ∎

7.2.6 *Cellular switches*

The patterns of molecular density discussed in the previous two sections may describe the behavior of patterns of pigment. More generally, in developmental biology it is

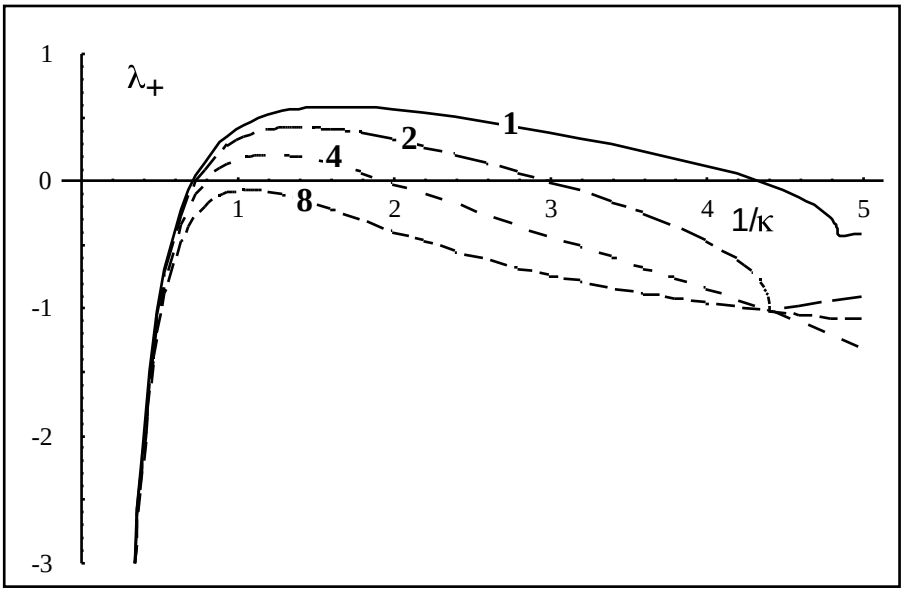


Figure 7.2.17 Plots of the real part of the eigenvalue λ_+ as a function of $1/\kappa$ as obtained in Question 7.2.8. The real part of λ_- is always negative for the parameters chosen. The plots are for parameter values: $D_a = 0.5$, $D_b = 20$, and $k_1 = 1$. The value of $k_3 = 1, 2, 4, 8$ is indicated on each curve. We see that the range of length scales over which the uniform solution is unstable decreases with increasing k_3 and eventually vanishes, causing the uniform solution to become stable at $k_3 = 8$. This is consistent with the simulations for $k_3 = 1, 2, 4$ shown in Figs. 7.2.14 and 7.2.15. The uniform solution (not shown in Figs. 7.2.14 and 7.2.15) was indeed found to be the result of simulations at $k_3 = 8$. ∎

necessary to use such patterns to activate certain cells to perform specific functions, change shape or initiate another stage of pattern formation. For any of these to occur, a chemical process inside a cell must be initiated. The chemical process should persist independent of the original pattern of molecular density. Then the pattern itself need not persist as the system further develops. This requires a one-way chemical switch that can then activate additional cellular functions.

In order to realize the behavior of a one-way switch, what is needed is a chemical system that has two stable states and can be switched from one to the other by a prespecified concentration of the patterned molecule. The prespecified concentration is genetically encoded to achieve the desired control. We require a new reaction equation that depends on the concentration a of the patterned substance A and controls the concentration c of a substance C:

$$\frac{dc(t)}{dt} = h(a,c) \tag{7.2.65}$$

$h(a,c)$ must have the property that as a function of c it can have at least two solutions c_{-1} and c_1 (in a moment we will see that it must have three) of

$$h(a,c) = 0 \tag{7.2.66}$$

which are the steady-state conditions in which c does not change. These solutions are functions of a, and can be assumed to vary smoothly with a. However, above a specifiable density of a, one of the two solutions,say c_{-1}, disappears. This causes the density of C to switch to c_1.

We can analyze the properties of $h(a,c)$ that are necessary and suggest specific forms it might take. In order for c_{-1} and c_1 to be stable solutions of Eq. (7.2.66), the derivative of $h(a,c)$ must be negative at these values:

$$\left.\frac{dh(a,c)}{dc}\right|_{c_{\pm 1}} < 0 \tag{7.2.67}$$

This means that a small positive increment results in a negative dc/dt (see Eq.(7.2.65)) while a small negative increment leads to a positive dc/dt. In either case $c = c_{\pm 1}$ is restored.

The burden of creating a switch is on the density c, so we represent simply the effect of A on C as direct production ($A \rightarrow C$), or catalysis of production ($A \rightarrow A+C$) leading to the form:

$$h(a,c) = k_1 a + \tilde{h}(c) \tag{7.2.68}$$

We can now design $\tilde{h}(c)$ with the desired properties and consider how it can be generated using chemical reactions. For simplicity, we set $\tilde{h}(c)$ to have its first steady state at $c_{-1} = 0$ so that there is no constant term in $\tilde{h}(c)$. Since it must have a negative derivative at $c_{-1} = 0$ we have $\tilde{h}(c) = -k_2 c + \ldots$ where the ellipsis represents higher-order terms.

In order to have two solutions of Eq.(7.2.66) with negative derivatives, $\tilde{h}(c)$ must have a form like that illustrated in Fig. 7.2.18(a). In particular, there must also be a

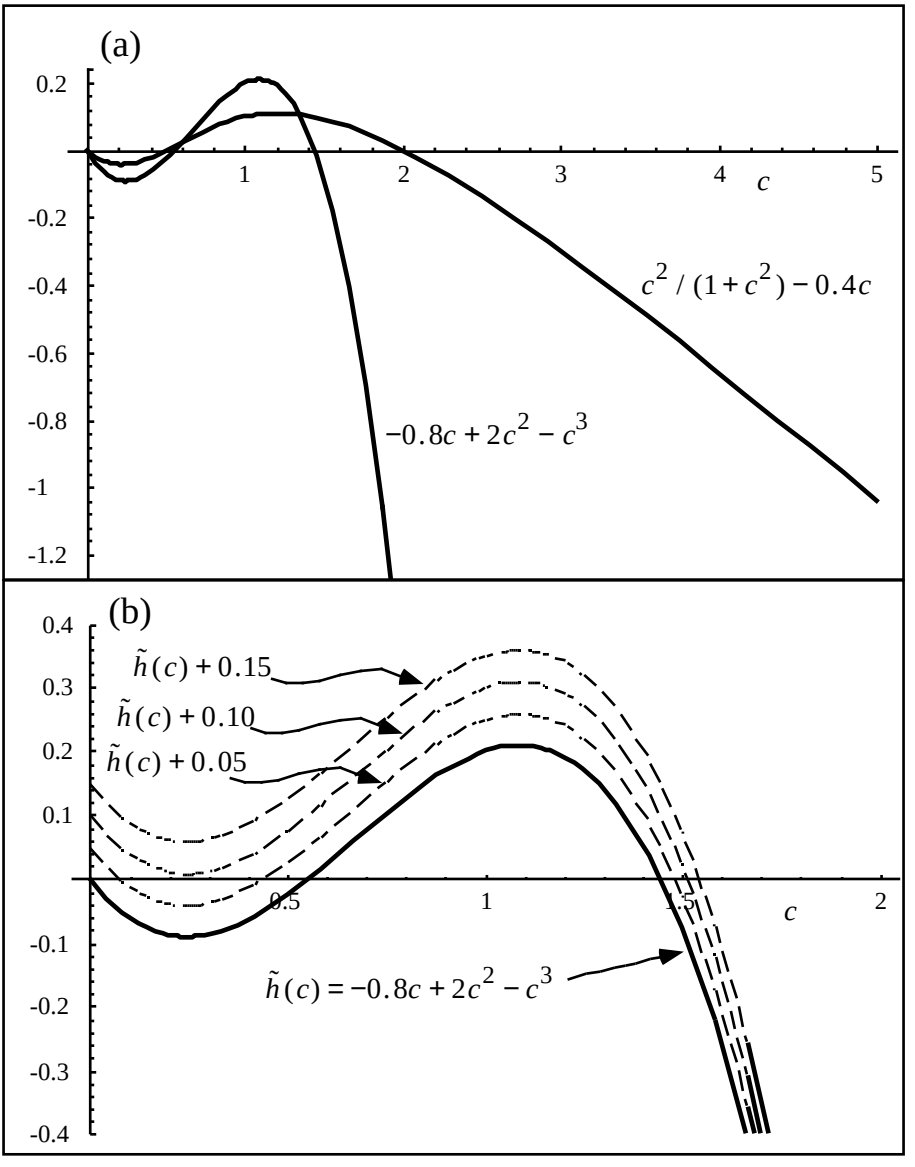


Figure 7.2.18 (a) Plots of two possible forms $\tilde{h}(c)$. This function describes the rate of change of the density c used to create a chemical switch. There are two stable solutions (low and high density) and one unstable solution of $\tilde{h}(c) = 0$. (b) When a is added to the system the curve is displaced upward, as shown by the dashed lines. For a high enough value of a only the high density solution is left. If we start with the low-density solution and raise the density of a the density of c will rise gradually and then switch to the high-density value. When a is lowered back down, c stays at the high-density solution. This sequence describes turning the switch ON. ∎

third solution of Eq. (7.2.66) with a positive derivative. This can be achieved using a polynomial of the form:

$$h(a,c) = k_1 a - k_2 c + k_3 c^2 - k_4 c^3 \tag{7.2.69}$$

For the last term, we can use any power of c that is greater than 2. Writing down reactions that in principle would lead to this form is not difficult. However, they may appear overly contrived.

Another way to satisfy the conditions is to make use of a system that has the structure:

$$h(a,c) = k_1 a - k_2 c + \frac{k_3 c^2}{1 + k_4 c^2} \tag{7.2.70}$$

The third term has the interpretation that it consists of a molecule produced in a sigmoidal fashion—it increases quadratically by autocatalysis and then saturates at a maximum value. In both Eq. (7.2.69) and Eq. (7.2.70) the second term represents a process of molecular decay. Fig. 7.2.18(b) shows the switching action when there is a change in the concentration of a.

7.2.7 *Pigment cell diffusion*

The study of the formation of patterns in Sections 7.2.2 and 7.2.5 considered systems where the initial conditions provided pigment placed at random throughout the space. The dynamics then caused these pigment molecules to bunch together to form the pattern. Experiment suggests, however, that vertebrates create pigment patterns by the migration of pigment-producing cells (melanophores). Early in fetal development, the melanophores are formed on the line that eventually becomes the spinal cord and from there migrate across the surface and aggregate into a pattern that becomes the pigment pattern. The number of these pigment-producing cells need not be conserved during this process, however, they must arise in most regions by migration, rather than by initial seeding or by spontaneously being produced by other cells.

Thus, we must consider a model where the initial conditions place pigment only in a limited part of the space, and from there the pigment diffuses through the space to form the pattern. We consider this process in the context of the reaction-diffusion systems described in Section 7.2.5. The slow diffusing species is the melanophore, while the fast species is assumed to be a molecule (Question 7.2.9). In both the activator-inhibitor and activator-substrate systems, the slow-diffusers (A) are not spontaneously generated—some of A is required in order to make more of A—consistent with the properties of melanophore reproduction. However, both of the models must be modified to allow the initial conditions to consist of only a single initial band of A and B (Fig. 7.2.19 (top)).

For the activator-inhibitor set of equations, the problem with the initial conditions arises in regions where b is zero. The first term in Eq.(7.2.45) diverges. This occurs not just because of the initial conditions but also because B is generated by the

presence of A, which is limited in space by our assumptions. Thus, as discussed in Question 7.2.5 (Eq. 7.2.38), we introduce an additional constant k_6:

$$f(a,b) = k_1 a^2/(b + k_6)(1 + k_5 a^2) - k_2 a$$
$$g(a,b) = k_3 a^2 - k_4 b \tag{7.2.71}$$

The results of simulations are not very sensitive to the value of k_6, which was chosen to be 0.1.

For the activator-substrate equations (Eq.(7.2.45)),the problem arises from the uncontrolled growth of B in the regions where A has not yet reached. This eventually causes the simulation to break down as the gradients in B become too large to be integrated using the parameters chosen. It makes sense to limit the spontaneous generation of B using an additional parameter k_6 in the following way:

$$f(a,b) = k_1 a^2 b/(1 + k_5 a^2) - k_2 a$$
$$g(a,b) = k_3(1 - k_6 b^2) - k_4 a^2 b/(1 + k_5 a^2) \tag{7.2.72}$$

A quadratic term rather than a linear term was used so that the first-order expansion of the function would not be affected. The first-order terms, as discussed in Questions 7.2.7 and 7.2.8, play an essential role in the existence of patterns, while the higher-order terms are less crucial. A value of $k_6 = 0.1$ was found to be reasonable and was used for the simulations. It limits the growth of b to no more than $\sqrt{10}$.

The simulations of these two systems are quite distinct. Simulations of the activator-inhibitor system are shown in Figs. 7.2.19 and 7.2.20. For certain values of the parameters, the pigment does not expand out of the region in which it started. This can be understood when we think about how this system functions. The pigment cells A produce the fast diffuser B which inhibits the formation of A. Since the highest concentration of B is in the immediate vicinity of high concentrations of A, it becomes difficult if not impossible for A to move into additional areas. For other values of the parameters,the initial line of pigment is unstable to bending, and the pigment expands to fill the space in spots or stripes, or combinations of spots and stripes. An example is shown in Fig. 7.2.20.

In contrast, pigment in the activator-substrate system (Figs. 7.2.21 and 7.2.22) generally expands to fill the space. This occurs because the fast diffuser B, which enables A to increase in concentration,is readily available in regions away from regions of high concentration of A. The melanophores A diffuse outward and increase in numbers due to the availability of B. It is helpful to recall that inhibition in the growth of A arises only when regions of high density of A surround a region of low density. In the central region, A cannot grow because the density of B is maintained at a low level due to reaction with the surrounding A.

One of the patterns that appears in these simulations are stripes that run parallel to each other. In the activator-inhibitor model (Fig. 7.2.20),they form by extension of each stripe and they are essentially perpendicular to the originating line (spine). In contrast, the stripes in the activator-substrate system (Fig. 7.2.22) are formed

sequentially and are parallel to the originating line. Depending on the parameters, the whole space may become stripes, or stripes may give way to spots.

Many animals have stripes that are better described by these results than by the patterns formed from random initial conditions of Sections 7.2.2 and 7.2.5. Some have stripes that run head to tail. These are more easily accounted for by the activator-substrate model. In particular, patterns with two stripes along the spine and dots below are found (e.g., genets) similar to Fig. 7.2.22. In other animals, such as the zebra, stripes run perpendicular to the spine. This could be generated by a version of the activator-inhibitor model where the stripes originate along the spine. Alternatively, if the pigment cells only originate at the skull, the activator-substrate model might form the stripes sequentially. We can identify which model is reasonable from the pattern by noting whether the stripes are continuous across the spine. In the activator-substrate model the stripes would be continuous across the spine, while in the activator-inhibitor model the stripes would be broken at the spine.

7.3 Developmental Tool Kit

In this chapter we have focused on the modeling of pattern formation as a fundamental aspect of developmental biology. In this section we briefly describe other processes that are important or essential for the process of development. The ability to cause these processes to occur provides a tool kit for the formation of organisms with functioning interdependent organs and physical structures designed for particular tasks.

The formation of physical structures including organs, limbs and tissues involves various processes that occur both inside and between cells that change the number, shape and location of cells. Growth in absolute size of the system occurs by cellular replication (growth and division). Once cells have differentiated in function due to patterning, diffusion or directed motion of cells in chemical gradients plays an important role in the relative location of cell types. Programmed cell death also plays a role in the formation of structures. Changes in external and internal structure of the organism also arise from changes in the structure of individual cells, particularly the cell membrane. Oriented adhesion of cells also results from cell membrane behavior. These processes involve changes at the molecular level in the cellular membrane and cytoplasm. They are developmental processes within the cell that contribute to the development of the whole organism. Among the physiological structures that are formed are spheroids, balls, membranes, tubes and branching systems. In some areas intercellular spaces also become filled with various excretions of cells to form support structures for the cells and the whole organism.

For the study of patterns in growth, the formation of treelike branching structures (Section 1.10.2) is particularly interesting. In plants, these include external structures—branches and roots. Internal branching structures occur in plants (veins in leaves) and animals (veins, nerves, air passages in lungs, and duct systems in certain organs). Most of these are multicell systems that may be formed by elongation of tubelike structures through cellular division and growth, then a periodic or occasional

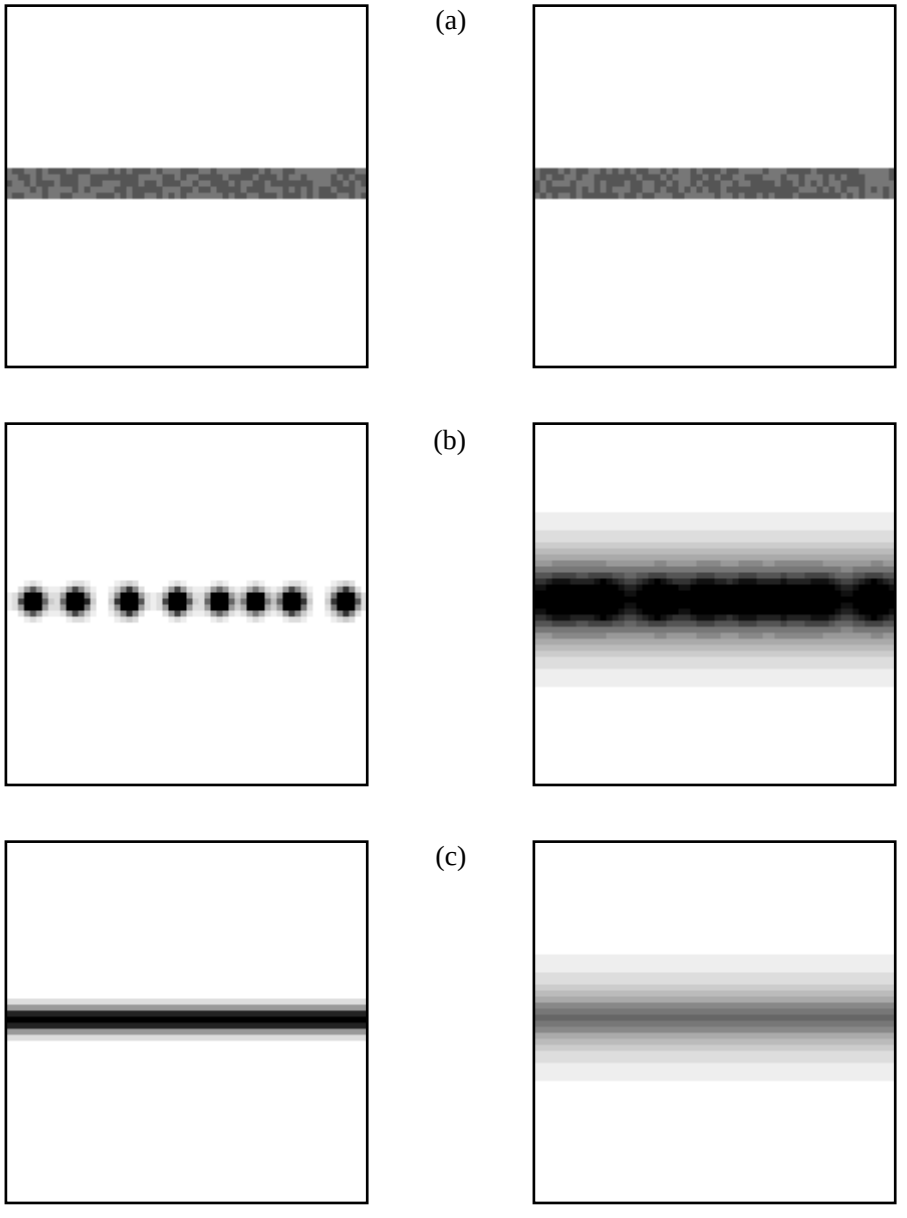


Figure 7.2.19 The reaction diffusion activator-inhibitor system simulated starting from initial conditions of a single linear strip of A (left frame) and B (right frame). (a) illustrates the initial conditions which are similar to that used in Fig. 7.2.13–7.2.16 but are restricted to a linear strip as shown. (b) shows the steady-state result of a simulation with parameter values $k_1 = k_2 = 1$, $k_3 = k_4 = 1$, $k_5 = 0$, and $k_6 = 0.1$. (c) shows a simulation with the same parameters except $k_5 = 0.1$. For the parameter values of both (b) and (c) the pigment remains confined to its initial line. ∎

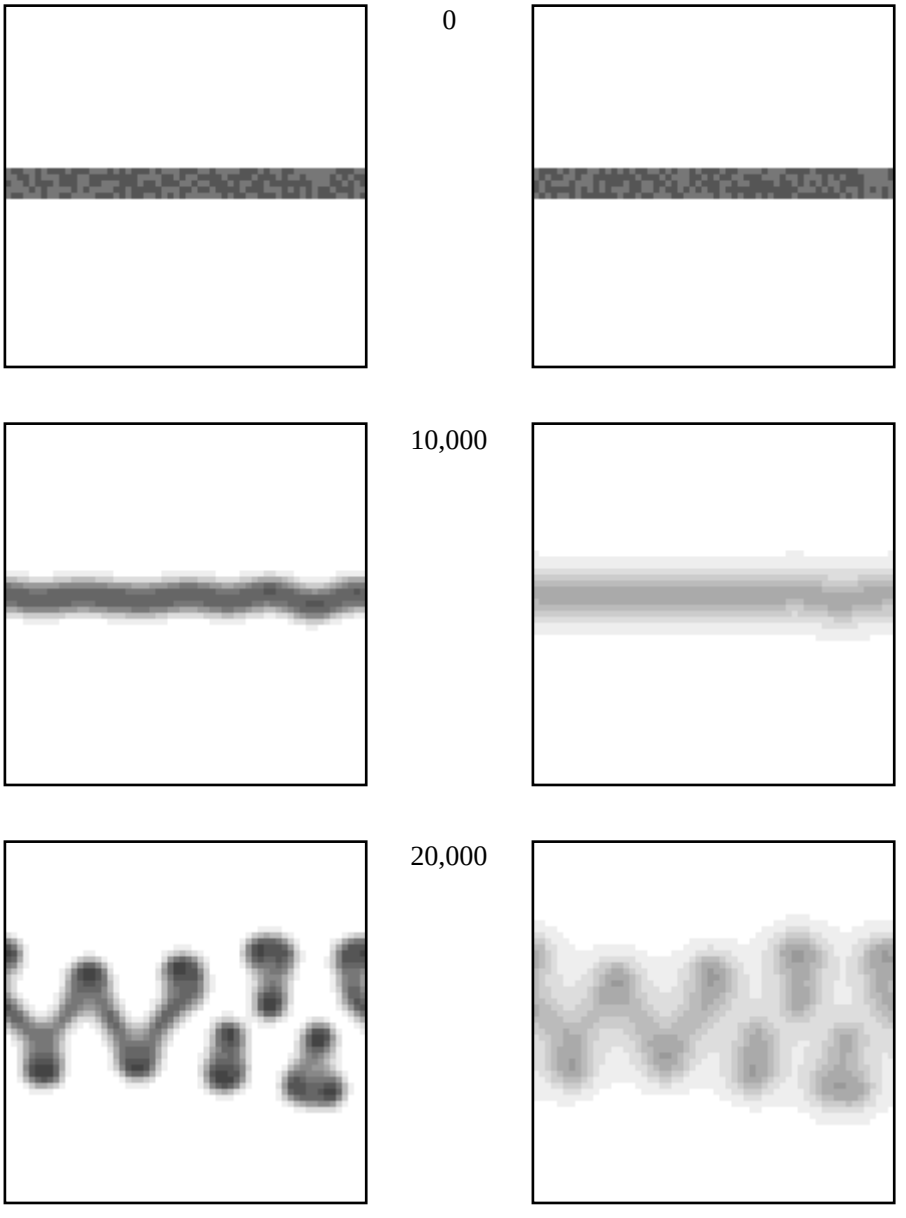


Figure 7.2.20 Frames of a simulation of the activator-inhibitor system with parameter values $k_1 = k_2 = 1$, $k_3 = k_4 = 2$, $k_5 = 0.3$, and $k_6 = 0.1$. The results are unlike the simulations shown in Fig. 7.2.19, for the same system but using different parameter values. In this case the initial line becomes unstable and the pattern of pigment expands to fill the space. Note that the lines of pigment are extended at their ends into the empty space. They are largely perpendicular to the line found in the initial conditions. Note also the long simulation time. The activator-substrate model has different behavior, as shown in Figs. 7.2.21 and 7.2.22. ∎

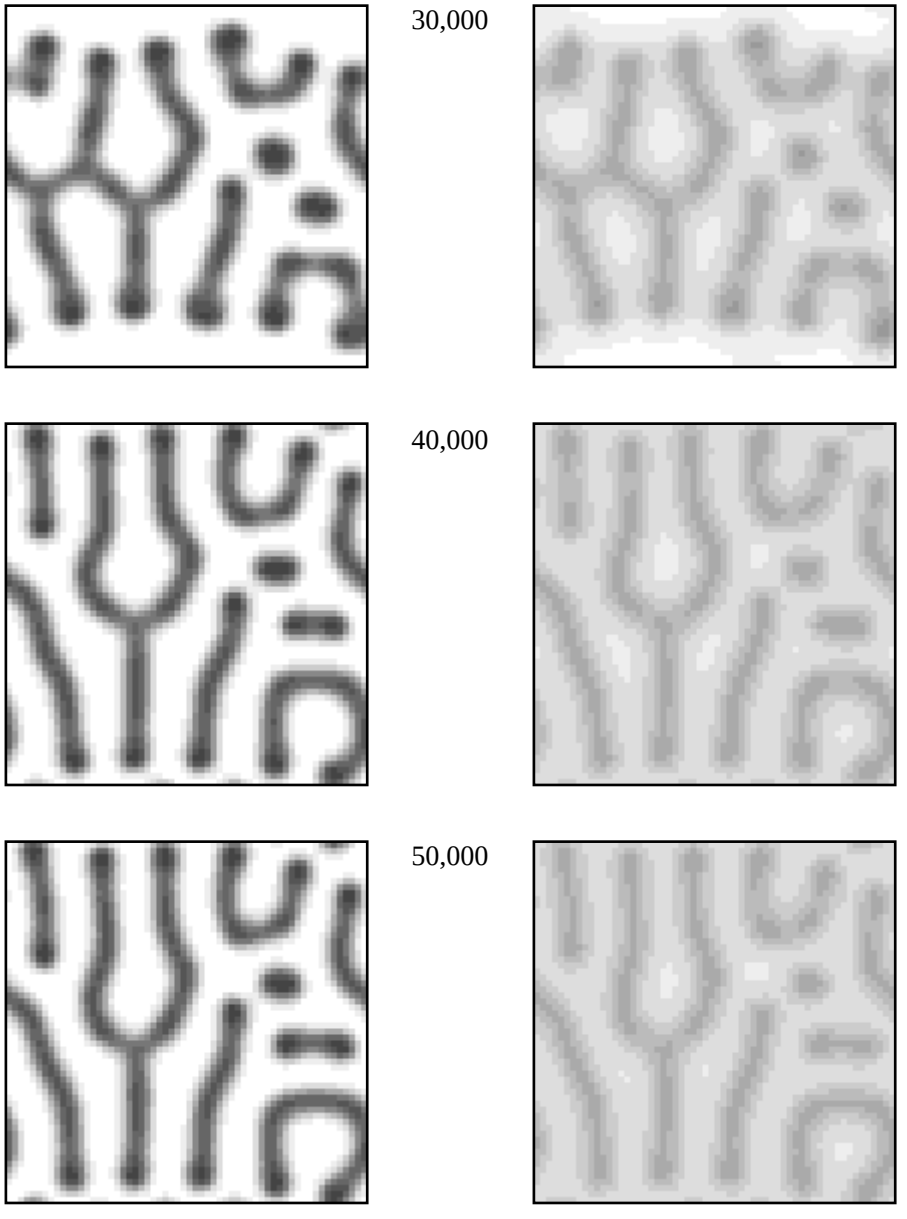


Figure 7.2.20 (*continued*)

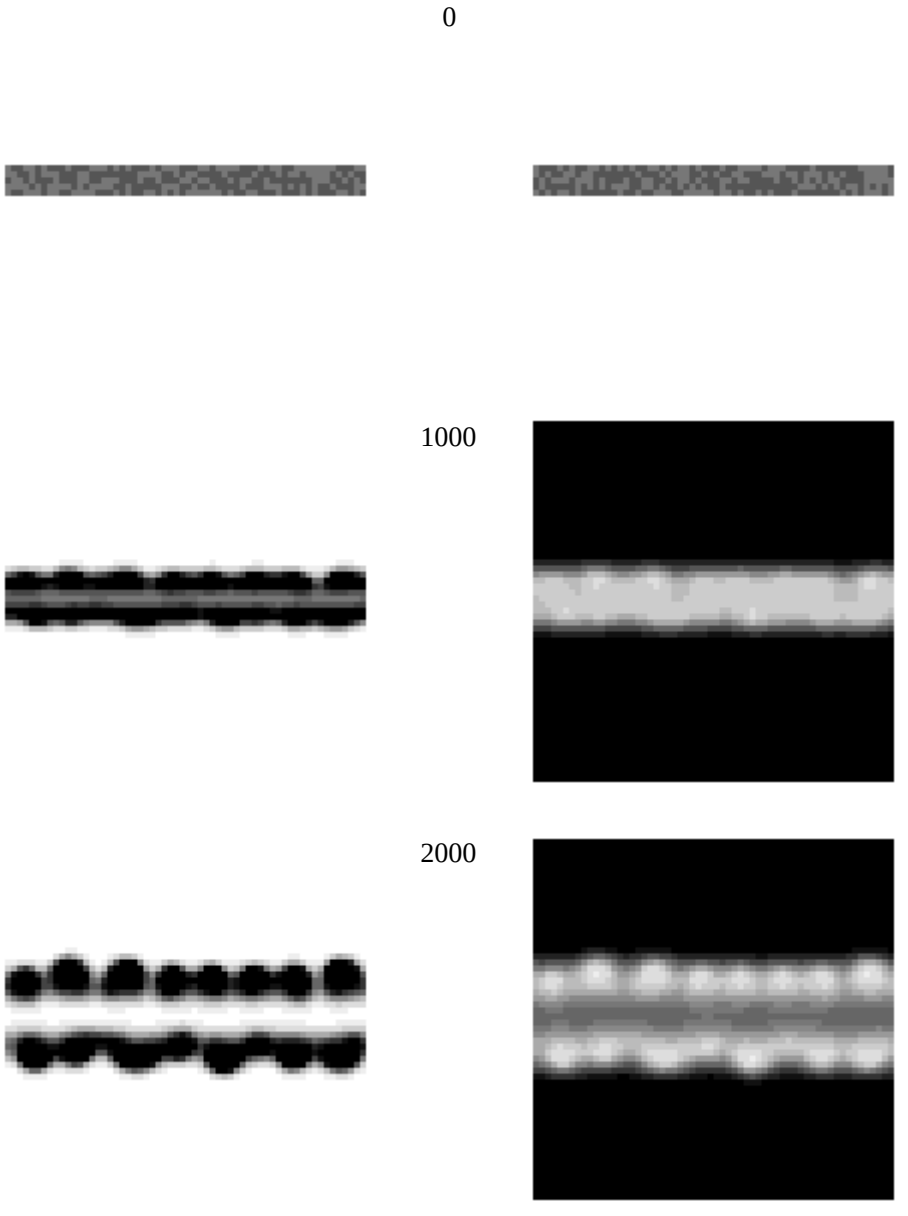


Figure 7.2.21 Frames of a simulation of the activator-substrate system with parameter values $k_1 = k_2 = 1$, $k_3 = k_4 = 2$, $k_5 = 0$, and $k_6 = 0.1$. The pigment expands to fill the space with spots using a process of spot splitting and diffusion. Compare Fig. 7.2.19 for the activator-inhibitor system. This model may also be relevant to evolution and trait divergence as discussed in Section 7.6. ∎

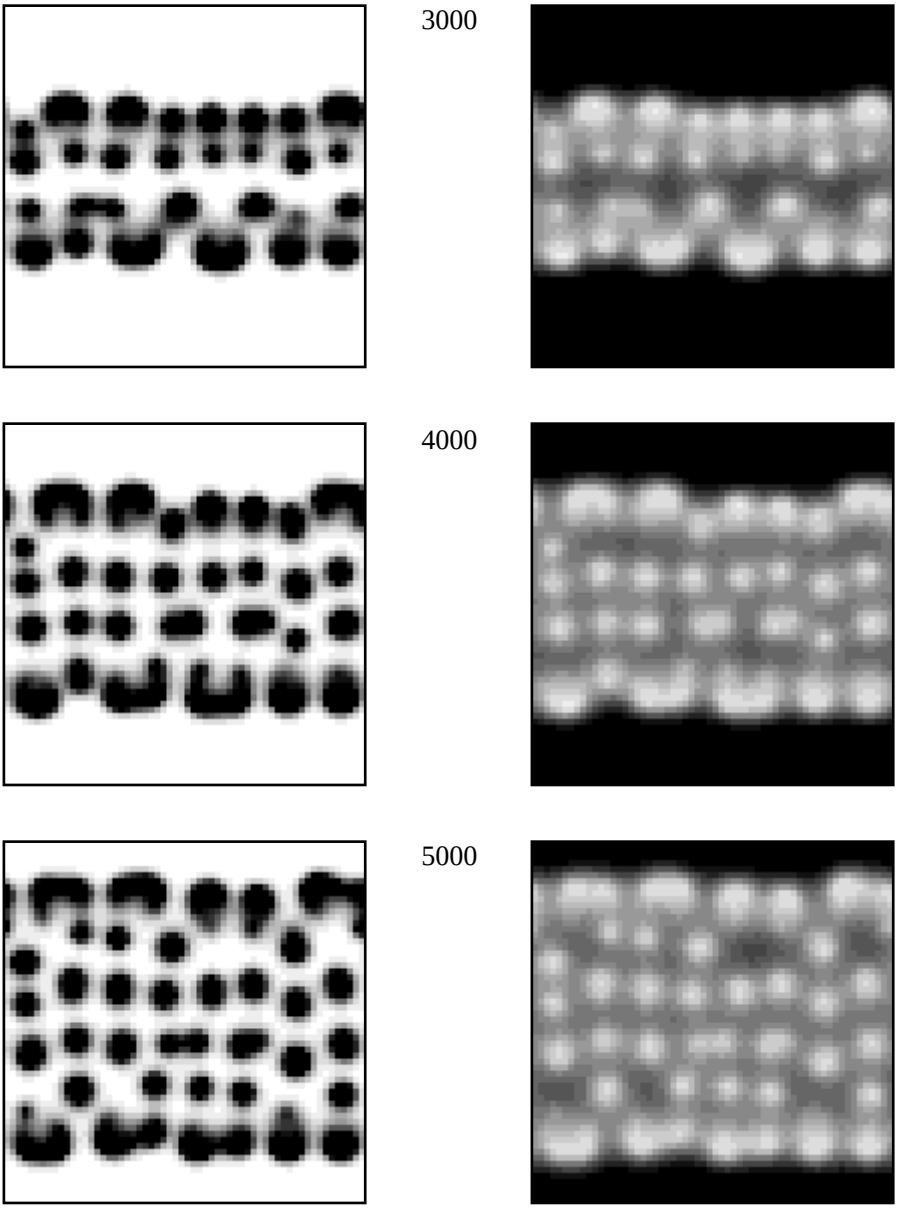


Figure 7.2.21 (*continued*)

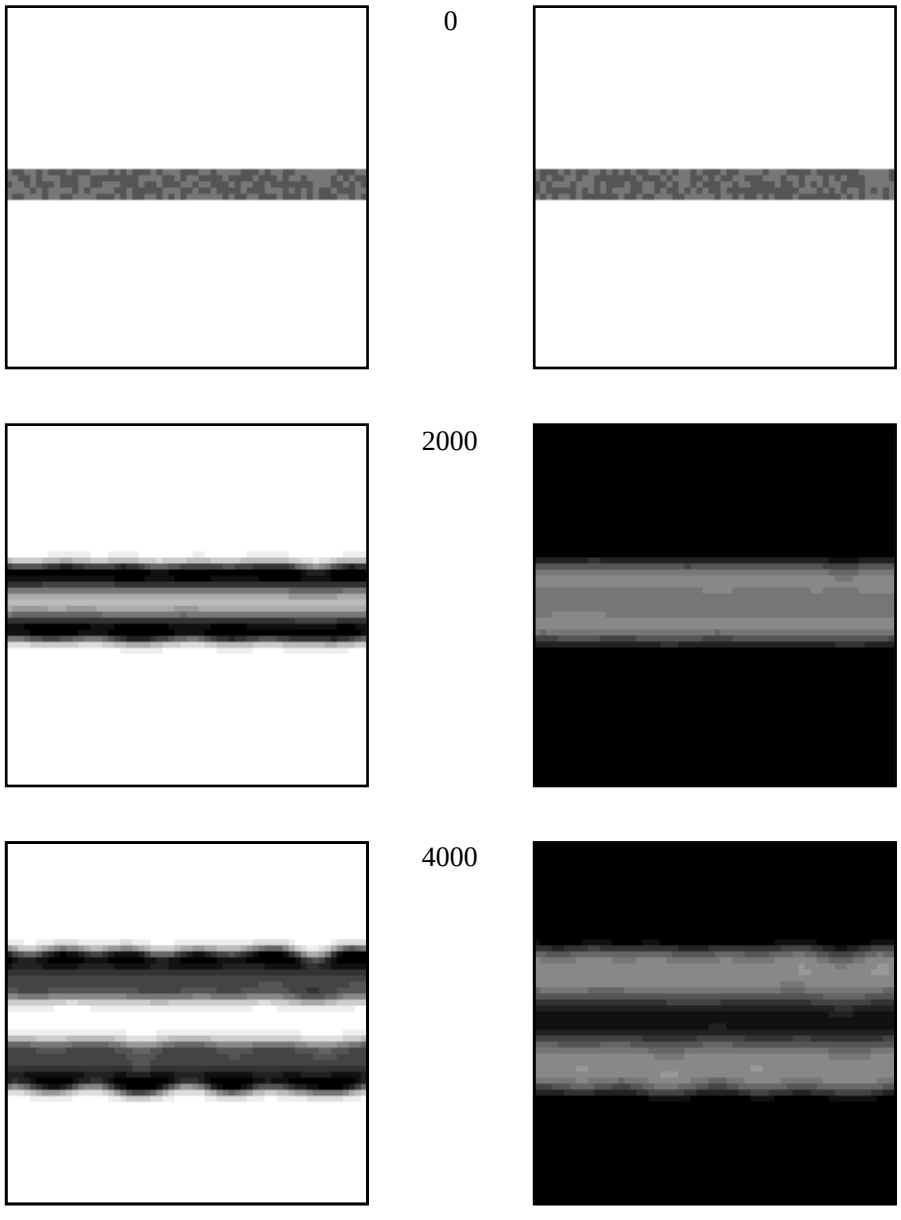


Figure 7.2.22 Similar to Fig. 7.2.21 except for the parameter $k_5 = 0.2$. We see that spots are formed when the original stripe diffuses outward. Then new stripes form parallel to the initial stripe of pigment by merging of the spots. For these parameter values, the spots continue to form into lines until the lines fill the whole space (not shown). All lines formed run parallel to the initial line of pigment. Compare Fig. 7.2.20 for the activator-inhibitor system. ∎

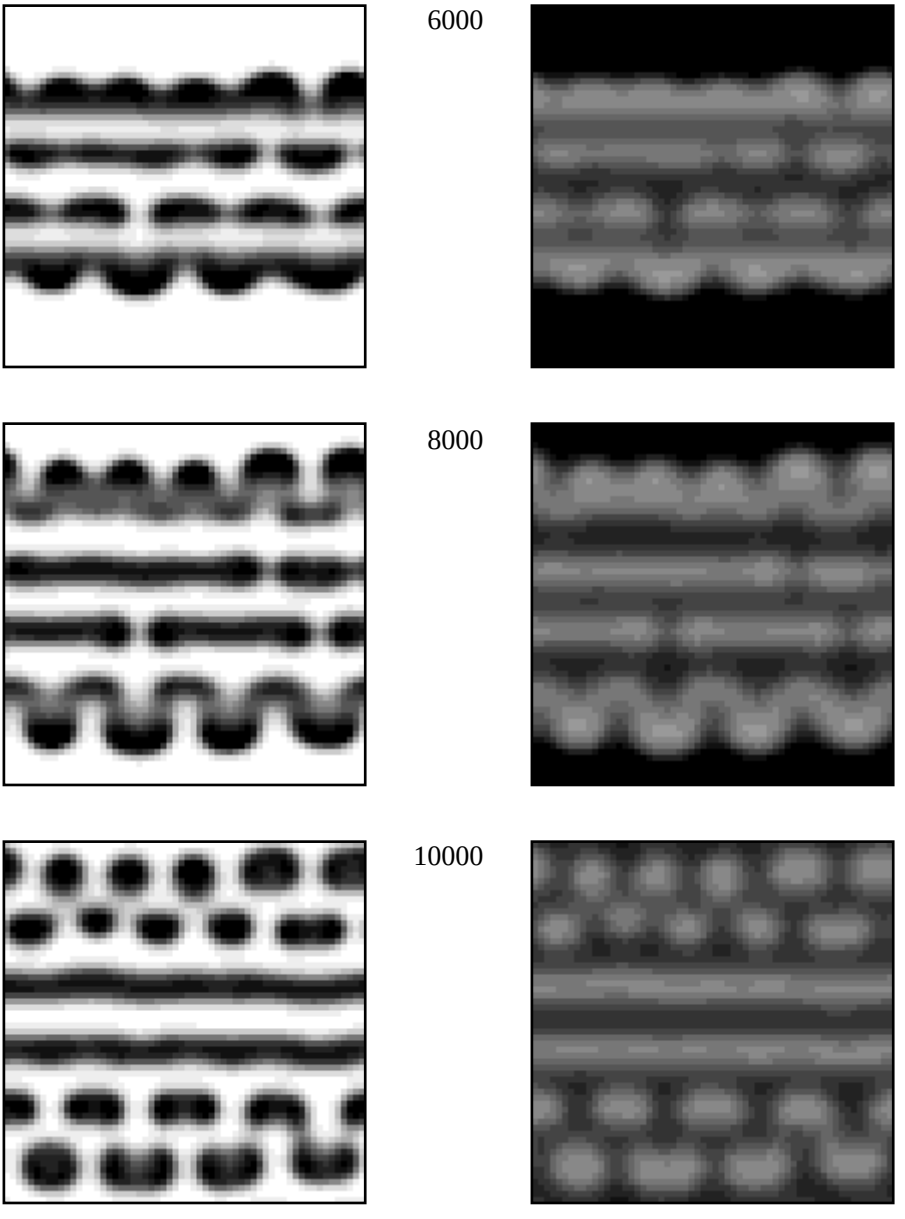


Figure 7.2.22 (*continued*)

initiation of branching. Some networks, such as the plant leaf veins, may also be formed by direct chemical patterning similar to the patterns shown in section 7.2. Branching nerve cells discussed in Chapter 2 are an example of individual cells that branch using molecular changes. Cell elongation and branching must be controlled through the addition of molecules to the cell membrane. Both multicellular and single-cell branching structures also, in general, must include a form of target tracking that imposes some overall systematic behavior on the branching system. This target tracking may cause the network to fill space more or less uniformly so that all regions are served (veins or ducts). Alternatively, a directional bias to growth may be important, such as provided by sunlight on tree branches, or chemical gradients causing a bias in growth direction that results in interconnection of organs inside an animal.

The mathematical modeling of branching structures would seem to be natural, since it is only necessary to specify an algorithm by which the branching occurs. However, there is a conceptual difficulty in representing such systems, because we generally think about the storage of information about a system in terms of storage locations that are themselves given by a linear string—a nonbranching data structure. When a branching structure grows, cells replicate at many sites, forming new cells whose existence and state must be specified. A better approach for describing tree structures, known as L-systems, has been developed by Lindenmayer based upon concepts originating in treelike hierarchies in linguistics. This approach uses a character string representation, but there are delimiters that indicate branching. Moreover, the dynamics allows the insertion of multiple characters at any site. These dynamics are specified by operators that act upon all the characters in a string. Each character can be considered as representing the state of a particular cell. We will illustrate this using a very simple example of a tree-generating algorithm.

We assume there are three states of a cell indicated by *A*, *B* and *C*. The update algorithm is specified by state transitions of cells that include the possibility of replication to two cells. Branchings are indicated by delimiters (brackets). A simple state transition table is:

$$\begin{aligned} A &\rightarrow A \\ B &\rightarrow AC \\ C &\rightarrow [B][B] \end{aligned} \tag{7.3.1}$$

The first few updates of a string are as follows:

$$\begin{aligned} &[B] \\ &[AC] \\ &[A[B][B]] \\ &[A[AC][AC]] \\ &[A[A[B][B]][A[B][B]]] \end{aligned} \tag{7.3.2}$$

This illustrates the representation of a tree with binary branches. The nongrowing part of the tree are cells in the state *A*. Cells in the state *B* replicate to extend the length of their branch, and cells in state *C* replicate to form two new branches. By further

elaborating such an algorithm,it is possible to specify geometric information that can fully describe a treelike structure. Various natural structures have been modeled in this way.

The formation of limbs through budding (including arms,legs,tail,head and fingers) may appear to be similar in many ways to the formation of branching structures. It may also be related to the formation of pigment patterns that specify the location of limbs to be formed. However, there is an essential difference between limb formation in animals and the other types of patterns. Both color patterns and branching structures can be treated primarily as a statistical process that allows significant variation between specific realizations. However, limb budding must be a reproducible process with definite outcome so that the number and type of limbs is consistent and directly controlled by the developmental process. Small-scale patterns involving only a few limbs would be much more reproducible and controllable than the large number of spots in patterns discussed in Section 7.2. The precise form of small-scale patterns is controlled by the boundary conditions that are imposed through the size (or number of cells) of the organism or the internal system in which the pattern is being formed.Our discussion in Chapter 2 of the 7±2 rule may be relevant here as well, suggesting a limit to the number of limbs that can be created reliably through such patterns.

Strictly repeating patterns such as faceted eyes of certain insects are another class of patterns that are different from those discussed in Section 7.2. Such patterns can be formed by sequential addition of elements. It is less reasonable to use a chemical patterning as a template to achieve strictly periodic order extending over a large number of elements. The main difference between periodic patterns and limb budding is not that there are few or many, but that in limb budding there are differences between the limbs that are important and the total number is well defined, while in a periodic structure all of the components are essentially the same and a few less or more doesn't really matter.

There are several other processes in addition to pattern formation and physical changes in cells that are important. These processes control the timing or order of developmental stages. We have already discussed in Section 7.2.6 the operation of one-way chemical switches that can serve to couple different processes. The presence of one chemical density above a threshold causes a second chemical to be produced. The second chemical continues to be produced even when the first is removed. Irreversible processes like chemical switches are an important component of timing mechanisms that count regularly spaced events. Timing mechanisms may be used for processes within a cell, including setting the time between cell divisions. Timing mechanisms are also necessary across several cell divisions. For this purpose, one way to monitor time is to count the number of cell divisions. This would require a sequential process (such as chemical switches) that can serve as a counter. It is believed that a certain number of bases at one end of DNA chains do not replicate in normal cell division (telomere shortening). The bases may be added later; alternatively, the progressive shortening of the chain may serve as a counter of cell divisions to control development and aging.

Our discussion of cellular processes in developmental biology is far from comprehensive,though a few of the important processes have been mentioned.A further level of detail could be added to the internal cellular processes. This level would include: the transmission of signals through membranes via cellular receptors, the transfer of such signals from the cellular membrane to the cell nucleus, the coupling of chemical processes to the activity of gene expression,the production of various enzymes,and the transmission of signals from their production sites out of the cell and into the intercellular fluid. Discussions of these processes are relevant to considering the cell as a complex system in its own right.

7.4 Theory, Mathematical Modeling and Biology

We have used various techniques to model pattern formation in biological systems. The primary tools were simulations of CA and differential equations. There is a need to develop some perspective on the utility of mathematical models for the study of biological systems. Biology is largely a phenomenological science. It is dominated by the experimental study of systems,their description and classification. This is to be contrasted with the physical sciences, where theory and mathematical modeling play a more integral role. At least for some biologists,the use of mathematical models misses the essence of the study of biological systems. Aside from the usual political/sociological issues that can affect such perspectives, there is validity to the concern that mathematical modeling may not capture the processes that are important in biological systems. It is important, therefore, to understand more systematically the objectives of theory in general and mathematical modeling in particular.

The role of theory in science rests on three legs—description, explanation and prediction. Description implies that a theory has the ability to describe the existing observations and phenomena. Explanation implies that the theory has a comparatively simple set of concepts and relationships that capture the system behavior more concisely than the phenomena that are described. This is tied to the notion of simplicity of scientific theory and Occam's razor, which requires a theory to be as simple as possible. Prediction is linked to the ability to describe existing phenomena but demands that clear and testable predictions be possible. In particular, a theory is considered poor if it cannot be falsified by direct experimental test. In essence the theory must distinguish between possible outcomes of an experiment whose implications would not otherwise be known. In a certain sense, the more unanticipated (surprising) are its predictions, the more useful is the theory.

From the point of view of experimentalists interested in further elucidating the phenomena of biological systems,the most important role of theory is the suggestion and prediction of the results of experiments. Indeed, every experiment that is performed is based upon some concept of what phenomena are important to measure, and therefore reflects a conceptual theoretical framework in which the experiment is performed. In biology, much of this theoretical framework is not based upon quantitative theory. As a consequence,there has been little expectation that significant quantitative predictions are possible. Recent efforts have demonstrated that constructive and predictive theories are possible, and the role of theory in biology is expanding.

In order to clarify the role of theory further, it is important to distinguish it from that of experiment. Experiment has a responsibility to uncover truths in the measurement of actual systems. Theorists are often assumed to have the role of proving truths through inference. However, their actual role is to propose assumptions—the theory itself—and correctly derive from these assumptions various predictions. Only when experiment tests the predictions can the assumptions themselves be tested and truth be determined. Because of the different objectives of theory and experiment, it is not appropriate to evaluate the contribution of theory by the goals of experiment. This is just as true about the evaluation of experiment by its contribution to the goals of theory. For example, in most cases experiment does not provide a general understanding, only an understanding of specific phenomena.

Increasingly, two additional roles of theory have arisen that cause more confusion about its ultimate responsibility. The first of these is the appearance of *ab-initio* calculations of system properties. This approach is most often applied to the study of solid or molecular systems. These studies extend the traditional objective of providing theoretical predictions for experimental results. However, the assumption in these studies is that the underlying theory has been so fully tested that the result of a proper calculation is as correct as an experiment that is performed on the same system. The challenge for the theorist is to ensure that the calculation is correct, if this is satisfied then the results are assumed valid. In this way it is like an experiment. The concept of *ab-initio* calculations has limitations in that there are no calculations that have perfect accuracy, and their implementation always requires assumptions about the relationship of the computer model with actual systems. Such limitations also apply to laboratory experiments and the relationship of the experimental condition to other circumstances. The objective of developing *ab-initio* methodologies is a positive one. However, it should not be confused with the more traditional objective of proposing fundamental simplifications and their experimental consequences.

The second additional role of theory is the mathematical modeling of experimental phenomena. This is known as phenomenological theory and represents a significant part of theoretical work in biology as well as in the physical sciences. Much of this chapter is rooted in phenomenological theory. While such theory is generally strong on description, it is weak on explanation and prediction. The reason for these difficulties is that a particular observation may be described by many distinct phenomenological theories. Thus we have seen that color patterns can be obtained from several different sets of differential equations and from CA. The general term for this feature of modeling is universality. The concept of universality implies that in many systems only a few aspects of the properties of a system are important in determining its characteristic (simple) behavior. The reason for this should be apparent: a simple behavior arising from a complex system cannot depend on all of the properties of the complex system. If it did, it would in turn be complex. Thus only a few features of the underlying system must be relevant, and many models should give rise to the same behavior.

A phenomenological model can be expanded to a more complete theoretical effort in order to provide additional information. One possible approach is to study directly the universality of the models. This means that we develop an understanding of the essential properties of models that can give rise to a particular phenomenology.

This approach takes us beyond the particular model and toward a general framework (theory) that provides a more systematic understanding of the origins of a phenomenon.One step in this direction is the articulation in this chapter of the principles of (a) activation and inhibition, and (b) fast and slow diffusers. Another is the analytic expansion of the differential equations in Questions 7.2.10 and 7.2.11. More formal approaches to universality are also possible (see Section 1.10.5).

A second possible approach is to discuss distinctions between different phenomenological models in order to provide contrasting predictions that can then be tested by experiment. This enables the phenomenological model to become more predictive and suggest experiments that can increase our understanding of the underlying causes of a phenomenon. The underlying causes themselves may not be readily accessible to experiment. For example, the discussion of diffusing melanophores and the difference between the activator-inhibitor and activator-substrate models in Section 7.2.7 provides a mechanism for distinguishing between the two models of pattern formation without direct knowledge of the actual processes involved. Without such a discussion there would be no way to tell which of the models applied for a particular animal except to study the molecular processes, and little would be gained from the theory. In general,the more independent tests are performed on a phenomenological model (or theory),the more it can be relied upon to describe new circumstances.

Most important for the consideration of the success or failure of theoretical modeling in biology is the recognition that complex phenomena require, by their nature, a complex model to generate them. This means that we cannot expect simple models to generate truly complex behavior. Thus, a basic skepticism about the ability of theory to describe biological phenomena can be justified. What is missing, however, is an ability to know, a priori, what are truly complex phenomena and what properties of complex organisms can be attributed to simple universal behaviors. Through a number of examples in this text, various approaches to the description of aspects and attributes of complex systems have been illustrated using relatively simple concepts and models. This is ultimately an important objective of the field of complex systems.

We conclude this section with a discussion of the relative utility of the CA, reaction-diffusion and other models of pattern formation as an illustration of the use of simulations in the study of biological systems. There are various biases regarding the use of particular forms of equations and this discussion is designed to illustrate that the form to be used should be dictated by the nature of the question that is to be addressed. We have seen that the CA models introduced in Section 7.2.2 were convenient for developing a basic understanding of activation and inhibition as a simplest model of pattern formation. The differential equation models in Section 7.2.5 provided a more microscopic view of these same processes in the sense that they modeled the chemical processes that might underlie the activation and inhibition. It would be important to recognize,however, that the particular differential equations used are not necessarily indicative of the actual processes in a biological system. They are thus only particular realizations of systems that embody the activation and inhibition phenomenon. These equations show us that reaction-diffusion systems can form pat-

terns. To achieve a yet more microscopic view of the processes, we might turn to another type of CA—the lattice gas—that would describe the diffusive motion of molecules directly rather than using the average density of the molecules as its essential variable. To be even more microscopic, we could use a particle model that includes Newtonian mechanics. This would require modeling the medium in which the particles are located. These examples are designed to suggest that there should be no inherent bias toward one approach. The bias is generated by the nature of the questions and answers that are desired.

7.5 Principles of Self-Organization as Organization by Design

In previous sections we discussed the dynamics of systems that achieve a complex structure through the interaction of their components. The context is our effort to understand developmental processes that are involved in the reproduction of multicellular biological organisms. There are, however, other processes that can result in reproduction. When a cell reproduces, it recreates itself by direct duplication. Each of the components is duplicated or simply divided in two parts and they are grouped to form two daughter cells. Does direct duplication play a role in the reproduction of multicellular organisms? Many plants can reproduce by growing new plants directly from a mature plant. Despite the connection to the parent plant, and its provision of nourishment, this is not duplication. Instead, differentiation and pattern formation occur in the creation of roots, stem and leaves. The other mode of reproduction, through a seed, is essentially independent of the parent plant. Thus the entire process is developmental. In animals, the interaction between parents and offspring is also secondary to the inherent developmental process. Fertilized eggs may be warmed by birds and the young may be fed and trained. Mammals have a more direct relationship, initially through a controlled uterine environment, then through nurture. Nevertheless, the specification of the process of physiological development is understood to be largely self-contained in the initial cell. It would be remarkable if it were found that some structures are transmitted directly from mother to fetus in-utero by migration of differentiated cells. However, the basic developmental phenomenology appears independent. The question that arises is, Why do biological multicellular organisms reproduce using a process of development? What benefit is there in this process?

In order to understand why a developmental process is desirable we should consider the general task of creating a complex system, and specifically the problems with duplication. The first aspect of duplication that we might consider is rooted in the difference between individual cells and multicellular organisms—the existence of more levels of structure. In order to duplicate a multicellular system, we would duplicate each cell and then we would have to disentangle the two resulting organisms. This problem is linked to the spatial structure of the organism. An essentially two-dimensional organism in three dimensions would not have this problem. Individual cells are able to overcome this problem when organelles in the cell are replicated and separated, though this is a complex process. By lining up DNA strands along a single

two-dimensional plane, this part of the system is reduced to two dimensions. We might consider whether there are ways to do this for multicellular organisms. Instead, we will look for other reasons for a developmental process that are also fundamental.

Another mechanism for duplicating a system would involve a process that is more akin to our manufacturing processes, where many copies of a system are produced. Note that for many multicellular organisms there actually is mass production of offspring, so that this is not an unreasonable model. Starting from a prototype or a description (representation) of the whole system, we create a process that produces and then places each of the components in its proper location. There are various problems with this for an interdependent complex system.One of these is that the system must be maintained in partial form. Sustaining the various components separately creates an additional burden on the manufacturing process. This problem exists in actual manufacturing, since structures that become self-supporting must be maintained during construction. Extrinsic supportive structures (scaffolding) may be necessary during construction that are later removed. For a very complex interdependent system such scaffolding would be much more difficult to design. Even for a developmental process, the problem exists. It is manifest in the support systems in a reptile or bird egg, and in a mammalian uterus. However, the internal organs are still largely maintained by self-consistent systems that develop into the mature systems of the multicellular organism.

While the two problems discussed in the previous paragraphs are important, there is another way to understand the reason for a developmental process, which will be particularly relevant for our understanding of the design of complex systems. It relates not to the structure itself but to the problem of specifying the structure. Any design process implicitly assumes that a description of the system exists before the system itself does. This description, generically called a blueprint, is in many ways like a model of the system. We can better appreciate how science and engineering are related when we recognize that the relationship between system and description plays an important though different role in both. Ultimately, it is the interplay between system and description that science is investigating. A key difference between science and engineering is that science can advance by using partial descriptions, while for engineering a useful description must be sufficiently complete.Our concern here is to understand the relevance of representation to developmental biology. In particular, What is the advantage of a representation that describes the developmental process of formation rather than the final system itself? This reformulation of our question suggests an answer: the developmental process can be more concisely described.

We can understand this answer when we think about the existence of various relationships between different parts of a complex system as well as the different activities of the parts. If we take advantage of these relationships, we can reduce the amount of information necessary to describe the whole system. More correctly, we use the relationship when we create the program of development that constructs the system. The program of development allows us to have less duplication of information. If the same basic structure is relevant to several components, we have them undergo the same developmental processes and then modify them later to accommodate dif-

ferences. Even after two components are different, the same developmental process may be used to achieve incremental modifications that are common to the two parts. The process that we are describing is the creation of an algorithm from which the system is to arise. The reason it is useful is because the explicit blueprint is inherently compressible. The algorithm appears to contain fewer pieces of information than the final form, even though they both ultimately contain the same information. We will enter further into a discussion of system representation and information theory in Chapter 8 (See also Section 1.8 and 1.9). The ideas articulated here are parallel to the idea of algorithmic complexity, where the notion of reducing the length of a description to its smallest possible representation (compressing a character string) is investigated. Here we are considering the applications of these ideas to design.

When we consider developmental biology in this context, we must expand our understanding of compression from the usual notion that allows only deterministic compression algorithms. Randomness or noise is available from molecular motion in biological systems. Information that is essentially arbitrary can be provided from this randomness. An example may be found in the pattern formation discussed in this chapter. To describe the patterns formed in all of their detail would require many pieces of information. However, for the animal skins, the specific details of the pattern are not essential—we can vary them and still have patterns with the same properties. As long as we are interested in the generic properties, such as size and overall shape of dots,then the details can be provided by randomness. In the simulations, this is provided either by the initial conditions or in the update process when there is a random selection of cells to update. To think about this more clearly we must recognize that the eventual state of the system is selected from an ensemble which results from the influence of randomness. As long as we are interested in properties that are generic in the ensemble,this is satisfactory. However, if we want to select a particular feature that is rare in the ensemble then we must specify it a priori as part of the design.

There is another source of information that may be used in the process of forming the system. This is the existence of specific well-defined influences of the environment. The environmental influences are in addition to the support structures and nutrition provided in the seed, egg or uterine development environment. We can illustrate this by another example. As mentioned in Chapter 3,the development of basic neural connections in the visual system of mammals is influenced by stimulation by light. This is not really a form of adaptation to the environment, it is instead the use of specific external stimuli as part of the developmental process. The algorithm is taking advantage of persistent information about the external environment—the existence of light.

We thus find that the process of development is convenient because it allows the system design to be more concisely represented. This answer is not complete. We must still explain why it is advantageous to have a more concise description of the complex system. The advantages of a concise design become particularly meaningful when the design is to be modified. Modifications should preserve many of the essential relationships encoded in the description. This reduces the possibility of design errors. In a complex system, a major source of errors is inconsistencies in the design. By

definition, an inconsistency reflects the violation of a constraint or relationship that is necessary in the final system. If the design automatically incorporates the required relationship in its compressed form, then the inconsistency cannot arise. More directly, if the description is shorter, there are fewer places errors can be made—the space of possible systems is reduced. This advantage is readily apparent when we consider the range of sizes and structures of mammals that contain similar internal organs with mutually consistent function and interconnections. It is also apparent when we consider the variety of cars that are produced and realize that systematic (algorithmic) relationships exist in the structure and placement of different components. A drawn blueprint cannot describe the interrelationships of engine size to car mass. Instead,the burden of applying such relationship is typically placed on human beings who know them as design rules.A developmental approach would incorporate these rules in an algorithm that could be modified to produce cars with various features and sizes. If the algorithmic description is sufficiently concise,then even random changes will still lead to viable designs. The importance of modification of design in biology is apparent in our discussions in Chapter 6 about sexual reproduction and the importance of random variation in evolution.

There are also disadvantages of a concise design.One arises when the design must be precisely duplicated. Without any redundancy in the information, copying errors may be introduced. We can see that the advantages of a developmental approach are most important when a design is to be modified frequently, and less so when it is to be duplicated many times without modification.Still,the problem of duplication can be largely overcome. This is done through compression with a limited number of redundancies that enable error detection.

Without further elaboration of these matters we can recognize the central issues that have been raised. The connections that we want to make in this discussion are between the biological developmental process, the design of complex systems, and the field of information and computation theory which is more commonly discussed in the context of computer algorithms.

There are various ways in which interrelationships or algorithms are used in the design of man-made products, whether these are physical entities such as cars and airplanes or computer programs.One common methodology is the use of modular design. For example, in the construction of apartment buildings or housing developments,identical units need not be individually described. If modules differ, however, they must generally be separately described. In order to execute the design, it is not sufficient to describe only the modifications. In software design, compilers or interpreters translate from a more concise, higher-level language into a form suitable for execution. We could also consider human elaboration of a design in a similar manner. The various stages in design development elaborate a concise specification. The first design might be an overall concept which is very concise. This concise description is elaborated in a process that can involve many human beings. Such a process is a kind of development when we include the human beings as part of the system. Thus we see that there are various ways in which algorithmic relationships are incorporated into

the design of man-made systems. However, it is apparent that the systematic use of algorithms is not yet well developed.

How can we further incorporate the concept of self-organization or algorithmic description in the design of man-made products? It is hard to imagine a developmental process that could create houses. However, it is not hard to imagine a computer-aided design system that can apply various modifications to a design and automatically incorporate design rules. Computer aided design in general can be understood as a process of elaboration of concise descriptions. In its present form it is not developmental in approach. Because the design description in a developmental process is more concise,the application of these concepts to design is a strategy for reducing the complexity of the design and engineering task. Our discussion of developmental biology suggests that it will become progressively important as the systems that are being designed become more complex and are modified more frequently.

7.6 Pattern Formation and Evolution

In Chapter 6 we considered various models of evolution as an undirected process that can give rise to complex systems. The essential concepts are the formation of diversity and selection from this diversity. In this chapter we considered models for pattern formation in developmental biology. The types of models we used are quite different. Here we point out that the mathematical models of pattern formation may also be relevant to the problem of evolution. Contact between these two problems arises from the pattern of organism populations in genomic or phenomic space and in physical space. The existence of a particular organism corresponds to a density $n(s)$ in this space. Species or trait groups correspond to patches of high density that are surrounded by regions that do not contain organisms. We assume that the pattern of populations is formed by evolution.

Evolution considered as diffusion in genomic space, including interactions, has resemblance to a reaction-diffusion system with some important modifications. We consider first the origin of short-range activation and long-range inhibition that may have given rise to the pattern of spots separated by unoccupied regions. This was already discussed in Chapter 6; we summarize here only a few points. There are various mechanisms for activation. The most direct is reproduction. Various social behaviors such as flocking are also mechanisms of short-range activation. Long-range inhibition must have a range that is reflected by the gaps between species. An important cause of inhibition is the consumption of resources. Similar organisms typically consume similar resources. Thus the existence of an organism causes inhibition of organisms over a range of genomes or phenomes. We can reasonably assume that the range of organisms that consume similar resources is larger than the range of organisms that are enhanced by reproduction. There may also be even longer-range interactions, but these we might include in a mean field treatment for the pattern-forming model.

With this motivation, we consider evolution modeled by a reaction-diffusion system with two components formed from the organism and its resource. The second set

of reaction-diffusion equations (Eq.(7.2.45)) is a natural model where the substrate b is the resource and the activator a is the organism. Organism diffusion is a consequence of mutation if s represents genomic or phenomic space. It is physical migration if s represents physical space. Resource diffusion need not be taken literally—the same effect (long-range inhibition) may be achieved due to organism behavior in consuming resources of various related types, or at various physical locations in the vicinity of its domicile.

For plant evolution we consider resources to be sun, water, nutrients and space. For herbivore evolution we consider plants and space to be the primary resources. For carnivore evolution we consider herbivores to be the primary resource. According to Eq. (7.2.45), the resource grows spontaneously but the organism reproduces when consuming the resource. The quadratic dependence of reproduction on organism population a in the terms $k_1 a^2 b$ and $k_4 a^2 b$ is a nonlinear or cooperative effect in consumption and reproduction. Sexual reproduction by itself only gives rise to a nonlinear dependence if the probability of mate encounter is small. If the probability of encounter is not small then reproduction is linear, since organisms are often limited to a certain number of offspring. The nonlinear dependence is suggestive of the cooperativity of effective consumption (e.g., a wolf pack or a lion pride) and resulting reproduction. The other term $k_2 a$ is the rate of organism death. From our studies of the behavior of reaction-diffusion systems,there are various modifications of this system that would still give rise to pattern formation, however, not all systems will result in patterns, and the pattern character varies.

There are two additional differences between an evolutionary model and the pattern-forming model: first, the existence of a fitness, or fitness-related parameters, that control the growth of population at a particular genome, and second, the existence of a higher-dimensional space than the two-dimensional space that we considered for patterns. To implement these modifications the equations would take the form:

$$\begin{aligned} f(a(s),b(s)) &= k_1(s)a(s)^2 b(s) - k_2(s)a(s) \\ g(a(s),b(s)) &= k_3(s) - k_4(s)a(s)^2 b(s) \end{aligned} \tag{7.6.1}$$

where we have just included the state dependence of all of the constants. They are all genome or phenome and location dependent, because the resources appropriate for a particular organism have their own dynamics, as do the organisms.

The essential behavior of this model without the species dependence of the parameters has already been simulated in the context of the pattern formation through diffusion of pigment cells. The modeling of diffusion of the pigment from a line in Section 7.2.7 is particularly relevant. We saw how patterns of spots can be formed that, in a model of evolution, would be interpreted as species or trait groups. The species closer to the starting line would correspond to simpler and more primitive organisms, while those far away would correspond to more complex organisms formed at a later stage of evolutionary history. We could readily imagine that such patterns will form in higher-dimensional spaces and with various species-dependent parameters. The degree to which variability of parameters would affect the relevance of such a model is still to be studied.

There are several advantages of a reaction-diffusion model for evolution that are appealing when contrasted with the models used in the previous chapter. The reaction-diffusion model gives insight into the reason that organisms continue to exist at different scales and at different stages of the evolutionary tree, including the coexistence of simple and complex organisms. We find this in the pattern-forming model, because the pattern continues to have populations in all regions of the space. The underlying reason for this is that the model inherently assumes that there is a variety of resources that are consumed by different organisms. A more complex organism that occurs later in evolution does not consume the same resources that a simpler organism does. To return to a model of the competition for a single resource, we would simply replace the many variables $b(s)$ with a single variable b. Or, more properly, we would expand the range of inhibition (by increasing D_b) to include the whole space. This would be similar to the renewable-resource model with only one resource. In this case, we have argued in Chapter 6 that only one type of organism would survive.

The model of a pattern-forming evolutionary process is also interesting in that competition is no longer the primary reason for the creation of complex organisms. Instead, the creation of complex organisms is due to the existence of resources that cannot be consumed by simple organisms. We might call these complex resources. Through mutation, organisms are formed that can consume the complex resources. Competition for resources causes the pattern of species or trait groups, but is not responsible for the existence of complex organisms.

We can modify this model to incorporate competition more fully by considering the space of resources and the space of organisms to be related in a more elaborate manner. Specifically, that organisms that are far apart in genome or phenome might consume the same resource. In order to know which organisms would be in competition, we consider the phenome space as projecting onto the resource space in a many-to-one map. As evolution proceeded, there would come instances in which organisms at different phenome locations but the same resource location would coexist, and the fitter organism would survive while the less fit would become extinct. However, there would still be a variety of resources giving rise to a variety of organisms at any stage of evolution.

In summary, the phenomenological existence of diverse species suggests that a reaction-diffusion model of pattern formation, with distinct resources for different organisms, is more realistic than a model that assumes a single resource for all organisms. The persistence of organisms over varied periods of evolutionary history and particularly the continued existence of organisms that originally appeared at much earlier stages of evolutionary history is suggestive of such a model. It is also consistent with the wide variety of resources found in nature.

The notion that the pattern of species is analogous to a developmental process of pattern formation also brings into focus the recognition that all life on earth is interrelated and in some sense is a single complex system. Loosely, by analogy, we might consider the collection of organisms on earth to be a collective organism similar to the collection of cells in a particular organism. This is relevant to the study of ecosystems and their behavior. We will address this from a more specific point of view in

Chapter 9 when we discuss the possibility that the collection of human beings on earth should be considered as a single complex system. This discussion will also have consequences for our understanding of the relationship between evolution and developmental biology. Before we do so we introduce and discuss in greater detail the concept of complexity in order to better evaluate the complexity of the global system of organisms on earth.Our focus, for various reasons, will be the global human civilization, but extending this discussion to include other organisms on earth is natural.