

The Struggle for Existence



Fall 2017

James D. Thomson

**BIO
120**

Ecology &
Evolutionary Biology
UNIVERSITY
OF TORONTO

The Struggle for Existence

Fall 2017

James D. Thomson

Citation: Thomson JD. 2017. The Struggle for Existence Fall 2017. Department of Ecology and Evolutionary Biology, University of Toronto.

Acknowledgements

Cover design: Laura Heslin Piper

Figure on page 14: Willem Swanenburgh and Jan van't Woudt from Wikimedia Commons

Figure 2: Ken Thomas from Wikimedia Commons

Figure 3: Copyright © 2018 by Hannah Fung

Figure 4: Copyright © 2018 by Hannah Fung

Figure 5: Copyright © 2018 by Hannah Fung

Figure 8: Allee WC, Park O, Emerson AE, Park T, Schmidt KP. 1949. Principles of Animal Ecology. Saunders, Philadelphia, USA. Copyright © 1949 by W. B. Saunders Company

The Struggle for Existence Fall 2017
Copyright © 2018 by James D. Thomson

All rights reserved. No part of this document may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author.

Table of Contents

Figures.....	1
Part 1 Introduction, motivation, goals, and advice.....	7
Part 2 Adaptation to meet ecological challenges.....	15
Part 3 Distribution and abundance as things we need to explain.....	21
Part 4 Patterns of climate from global to habitat scales.....	55
Part 5 Physiological ecology as a way of thinking about organisms.....	67
Part 6 Population ecology.....	87
Part 7 Community ecology, primarily regarding animals.....	121
Part 8 Community ecology, mostly plants.....	139
Part 9 Trophic relationships and plant-animal interactions.....	153
Part 10 Synthesis and conclusions.....	161
Literature Cited.....	167

Figures



Figure 1a. *Penstemon strictus*



Figure 1b. *Penstemon barbatus*

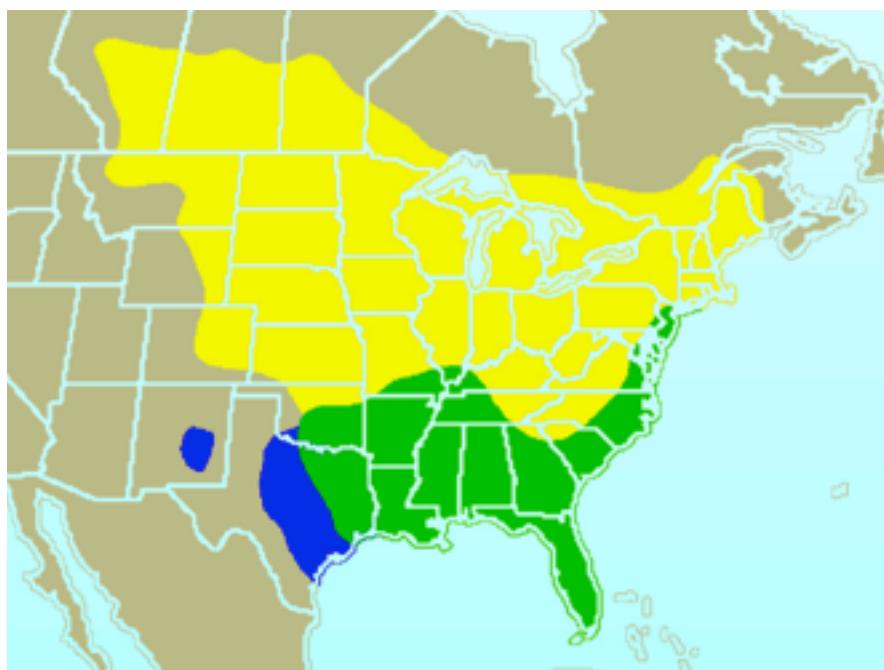


Figure 2. Map of winter (blue), summer (yellow), and year-round (green) range of the Brown Thrasher (*Toxostoma rufum*), a migratory bird with a large but not unlimited range. *Source:* Ken Thomas, Wikimedia Commons.

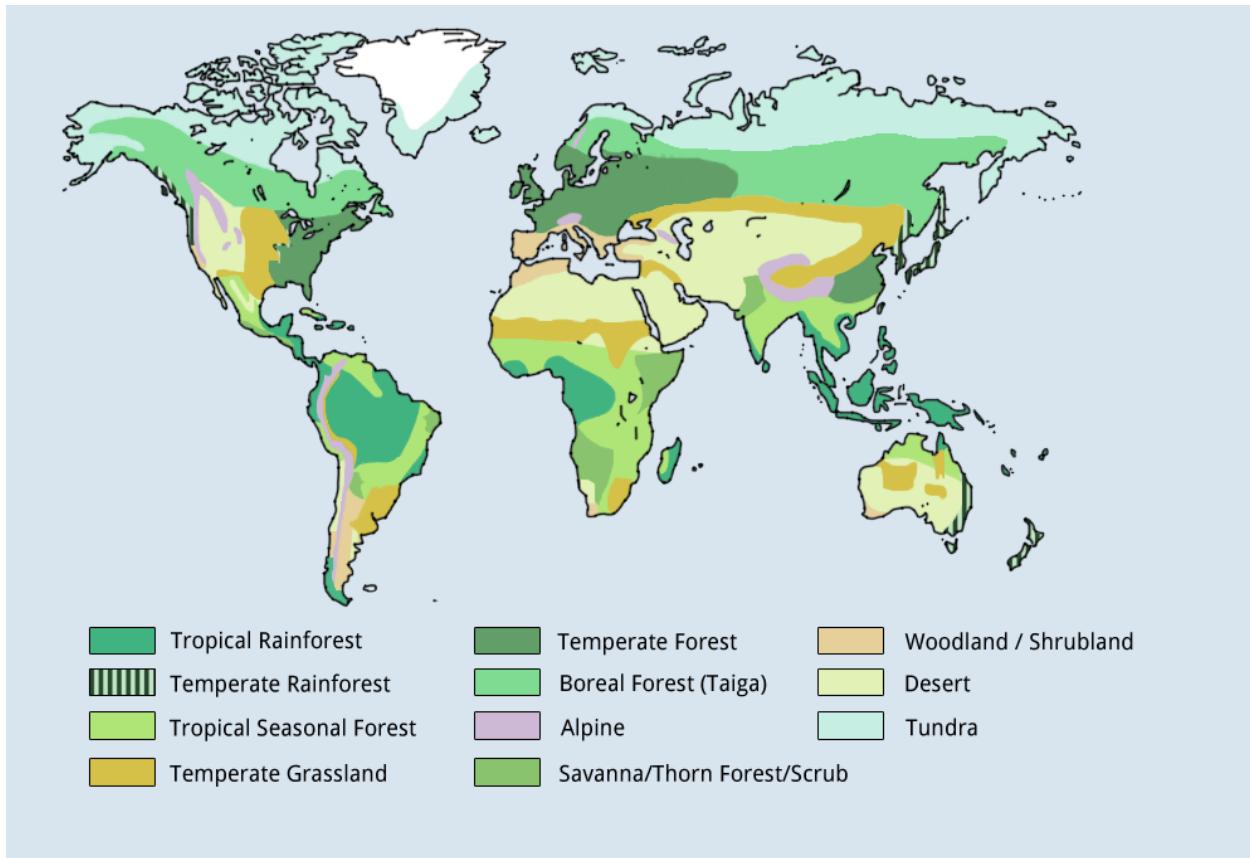


Figure 3. World biome map. This map of the major biomes of the world was prepared for BIO120 by Hannah Fung from several sources. You would be able to find many variants of it online. Different authors recognize different classifications of biomes, depending on whether they lump minor categories or split them out into separate types. Also, various names are used in different sources. The text explains some of those names and synonyms. This figure should be your primary reference for BIO120, however. Please study this map in close conjunction with Figure 4, which uses the same colour coding to illustrate what combinations of precipitation and temperature tend to produce which biome types.

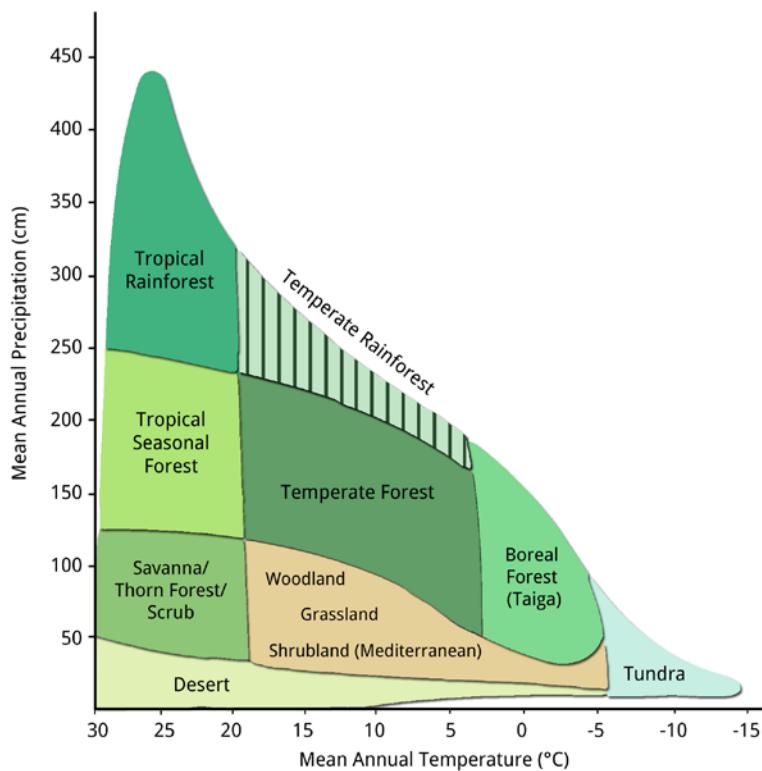


Figure 4. Whittaker's climate graph. This display shows how particular combinations of moisture and temperature tend to produce particular biomes. Drawn for BIO120 by Hannah Fung, it is based on a famous figure produced by R.H. Whittaker, who originally synthesized it from a broad literature on biogeographic patterns. It is meant to be studied in conjunction with Figure 3, which shows where in the world these various climatic regimes exist. As explained in the text, temperature and moisture are the two most important factors in determining the biology of a region, but other factors can also be important, especially the type of seasonality and whether the area has a continental climate or a maritime one. Therefore this graph has limitations, and it is impossible to produce an exact correspondence with Figure 3. For example, mountainous areas (shown in Figure 3 as "alpine") can't be portrayed as a sector on Figure 4's two axes, because both temperature and moisture vary so drastically with elevation. As discussed in lecture, a single mountain can host biomes ranging from desert through grasslands through forest to tundra. Also, the central sector of Figure 4 includes a mixture of biome types that have similar average annual temperatures and precipitation, but are differentiated by seasonality. See in Figure 4, for example, that "woodland/shrubland" and "temperate grassland" are found in different parts of the world, although their *average annual climates* are similar. In North America, for example, grasslands are found in the Midwest, while shrublands are represented by Mediterranean or chaparral habitats of southern California. What differentiates them is that the Midwest gets precipitation all year, with hot summers and very cold, snowy winters. Southern California gets almost all of its precipitation—as rain—during its mild, non-freezing winters. The Midwestern growing season (and its fire season) is the summer; the California growing season is winter and spring, and the fire season is summer. Those timing differences produce radical changes in vegetation, as explained in the text.

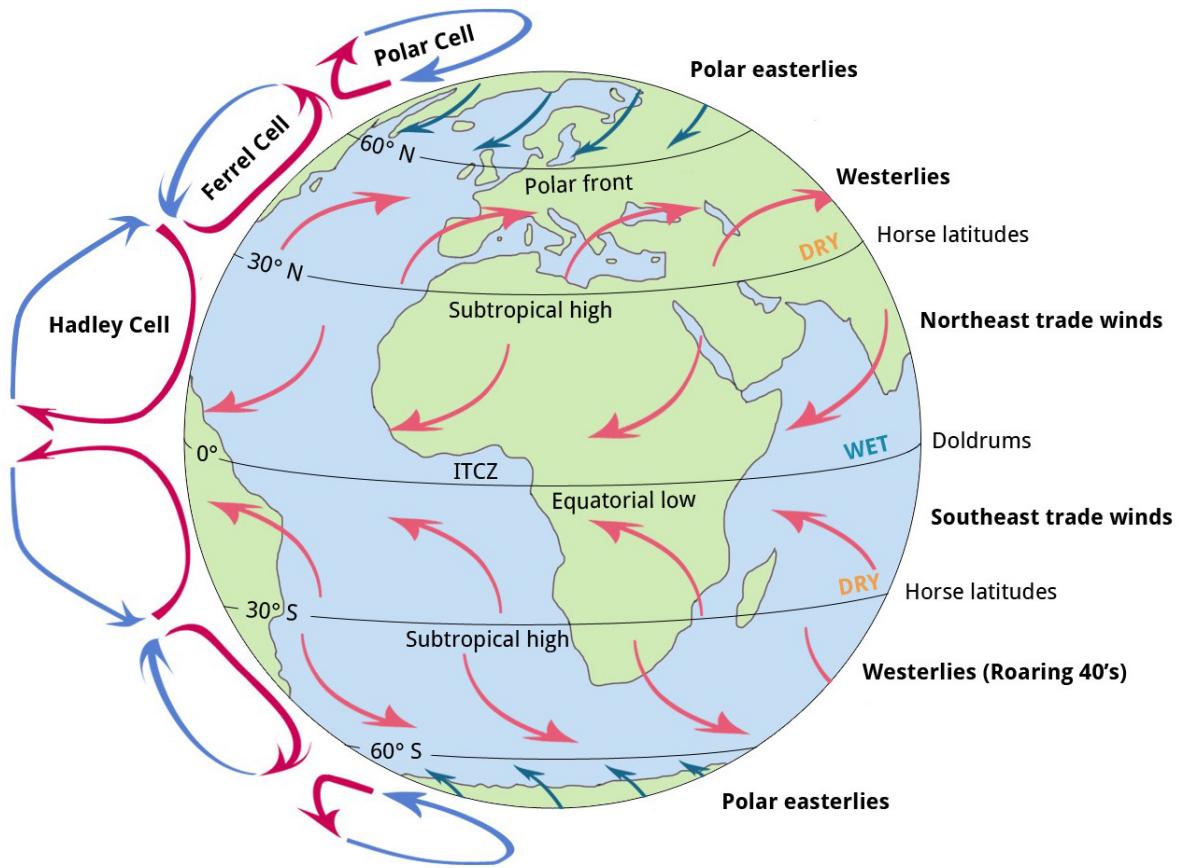


Figure 5. Global atmospheric circulation. Copyright © 2017 by Hannah Fung.

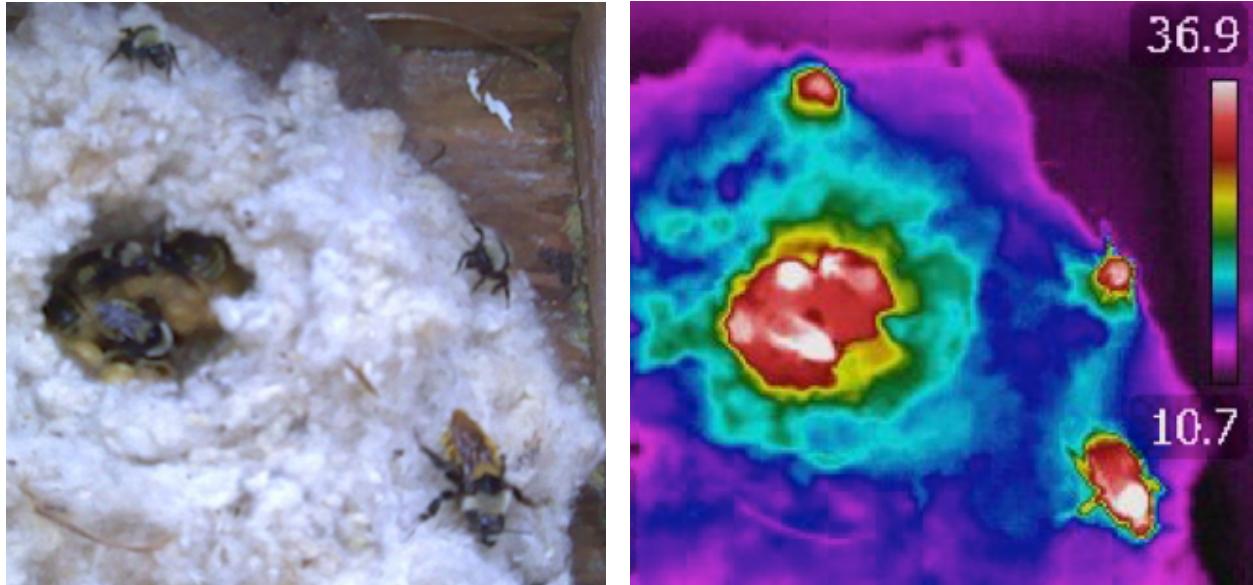


Figure 6. Bumble bees are furry, endothermic insects. Paired photos show visible light (left) and temperature (right) of a small colony of *Bombus appositus*, in an outdoor nestbox, photographed on a 7°C morning in Colorado, August 2014. The largest bee at lower right is the queen; the others are workers (her daughters). The bees have covered their brood clump with an insulating dome of white cotton, which we pulled back to reveal the brood clump, where all of the adult bees were incubating larvae (in wax cells) before we disturbed them. Note that (1) the larval cells are kept much warmer than ambient, (2) the bees get as hot as mammals do (37°C), and (3) their thoraxes are hotter than their abdomens, because they use their flight muscles to generate heat. *Photos by D. Inouye & J. Thomson.*

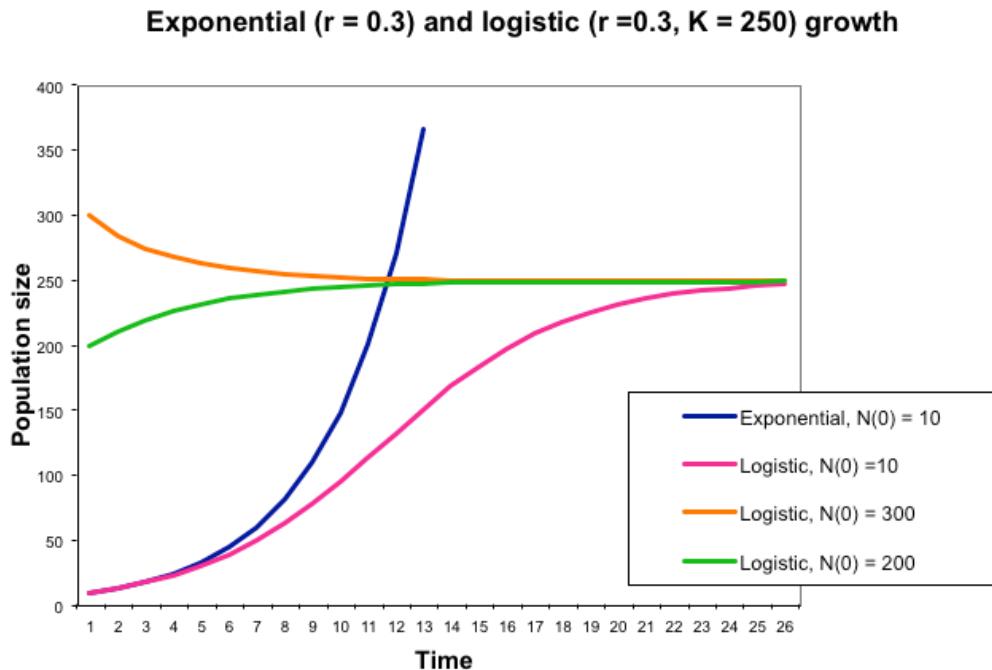


Figure 7. A population growing logically will always make a direct and smooth approach to the carrying capacity, from any starting point other than $N = 0$.

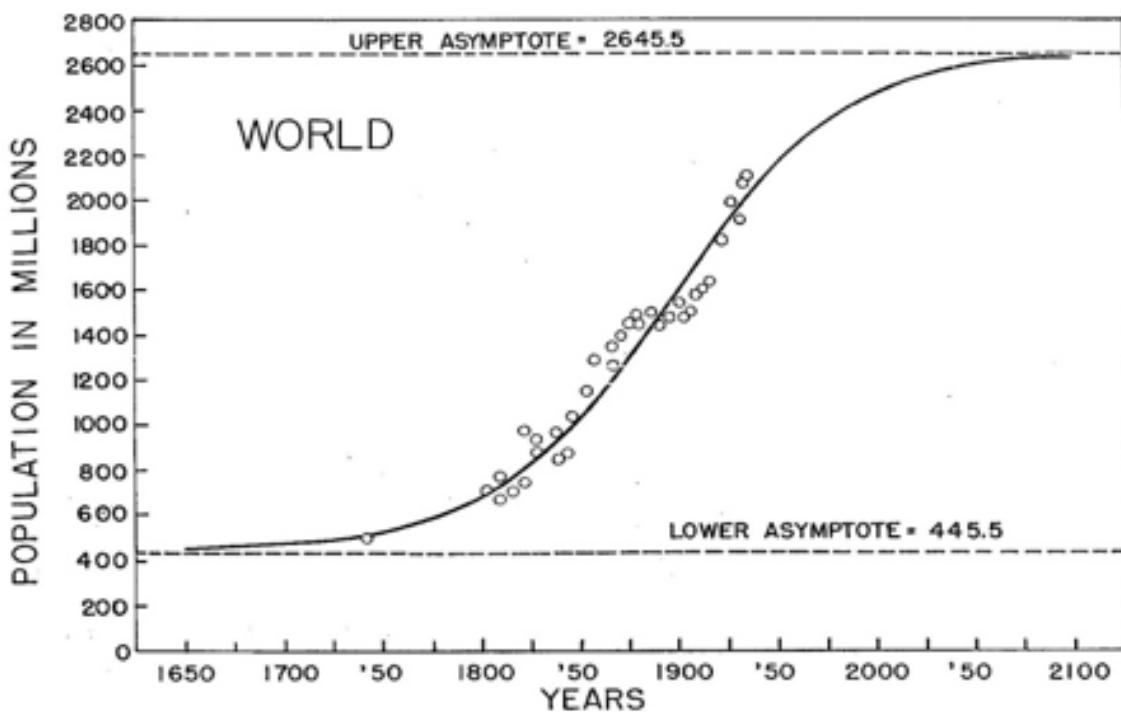


Fig. 107. The logistic curve fitted to an interval of recorded data from the population of the world. (From Pearl and Gould.)

Figure 8. The logistic growth curve fitted to human population estimates over time.
Source: Allee et al. (1949)

Part 1. Introduction, motivation, goals, and advice

Here we go: being more than a rock

Welcome to an unconventionally sequenced set of courses designed to set you up for various academic trajectories in the life sciences! Together, the Faculty of Arts and Sciences and the Faculty of Medicine at the University of Toronto offer a broad range of opportunities for increasingly customized study. To support that diversity of directions, we start with a very large course that tries to cover the most basic aspects of living organisms: their astonishing ability to meet the challenges of an unrelentingly hostile environment, to grow, to reproduce, to adapt, and to diversify. These abilities are so fundamental to our own lives as living organisms that we tend to take them for granted. To guard against that intellectual complacency, consider in how many ways an *organism*—you, for example—differs from an inanimate entity such as a rock. You could make up a long list. *The rock could not.*

BIO120 can be viewed as a survey of the major ways in which *life* on Earth differs from the lifeless minerals that make up Earth itself. This course's orientation deviates from the traditional emphasis in introductory biology. Traditional courses focus on surveys of the structural characteristics of the major groups of organisms. That approach is sometimes characterized as “the march of the phyla.” We consider such knowledge to be very worthwhile, but the expansion of molecular biology has put the squeeze on basic biology courses. Something has to give. Therefore, we concentrate on *fundamental principles and key examples* rather than exhaustive lists of characteristics. We defer the traditional surveys of organismic diversity and taxonomic organization to higher-level elective courses in EEB.

The *subject matter* in BIO120 is the diversification of life through adaptive evolution, driven by natural selection imposed by the environment. Our *approach* is to demonstrate *science as a way of thinking*. We separate the subject matter into an *ecology* segment and an *evolution* segment, although distinguishing those intertwined areas of study is rather artificial, and done mostly for convenience. I (James Thomson) wrote this short textbook to accompany my lectures for the ecology segment of the course. I am a professor in the Department of Ecology and Evolutionary Biology; my research interests center around the evolutionary ecology of plant-animal interactions, especially pollination.

A motivating question about sex, death, and you

In the popular imagination, high school biology is all about dissecting frogs. This cartoonish notion is probably outdated, but it doesn't matter; it was never what biology is about, even when high school students *did* dissect frogs. This is university: let's get serious.

Let's pick an arbitrary slice from the continuum of time. Roughly 200 years ago, there are exactly 256 special people spread across the world—most likely in Europe, Asia, and Africa. What makes them special? They are your great-great-great-great-great grandparents. Each of them carries an equal fraction of the genetic instructions that made you, and communally they carry the entire set. Fortunately for you, lust or love drives them into 128 sexual couplings. Then, certain of their children meet each other and do the same, generation after generation, with your genome-to-be aggregating into two-fold larger chunks at each step along the way. Finally, *you*, with your unique set of instructions, finally come into being through an utterly amazing process of sexual combination, fetal development, birth, and more development—all orchestrated by your particular genetic instructions. The zygote that will become *you* sets about harvesting and incorporating countless molecules. It organizes them into trillions of cells, laying down distinctive tissues and crafting organ systems of fiendishly interlocking complexity. Somehow, it all works. If you are a typical university student, you are now in robust, youthful health. Your recently matured brain is primed for brilliance, and you have great potential for a lusty reproductive career of your own. It is clear that life's processes have worked spectacularly well to make an improbable, incredible *you*.

But then it all unravels. Unless you die in some accident, your glowing health will gradually give way to sickness, senescence, and death. Somehow, what worked so unbelievably well doesn't work at all any more. This seems both paradoxical and unfair! Have you ever seriously tried to figure out a scientific explanation for *why you have to die?* (Yes, there are numerous faith-based explanations available—frequently contradictory—but in a science course, we search for evidence-based explanations.) *This* kind of big question is what biology is about. It's important.

What is this course about?

Suppose that the world contained only one species of organism, but that organism was able to live, reproduce, and die. It would be endlessly fascinating *for those attributes alone*. In fact, the world is home to an estimated 8.7 million species, give or take a few million. They *all* share those attributes, but they manifest them in different ways, despite having descended from a common ancestor. The most fundamental question we can ask about the world concerns this startling diversity: *why are there so many species of living things?* (Compare, for example, those 8.7 million biological species to the 98 naturally occurring chemical elements.) Since Charles Darwin and Alfred Wallace described the mechanism of evolutionary change by natural selection, scientists have come to

understand that the key concept is that organisms *adapt* to their environments—with the proviso that “environment” includes other organisms, not just the physico-chemical features of habitats. And because each organism is so thoroughly adapted to its circumstances, its morphology, physiology, and behaviour will be understandable only to the degree that we understand how evolution registers the stamp of the environment on the genome.

As its name implies, BIO120 concerns the phenomenon of *biodiversity* and the process of *adaptation*. We focus on these subjects using the powerful lens of evolutionary ecology. Every species represents a unique solution to the complex problem of living in a challenging environment. Why do we start here? Because *everything else in biology* can be considered a narrower subset, a special case, or an extension of these fundamentals. Our course concerns the broadest properties of life. This includes humans, of course, but BIO120 doesn’t place undue emphasis on human biology. Most of our examples concern metazoan animals (especially insects and vertebrates) and “higher plants” (angiosperms and gymnosperms) in terrestrial ecosystems. It is unfortunate to downplay organisms such as microorganisms and fungi. These organisms play critical roles in individual physiology, ecosystem processes, species interactions, and pathogenesis, but time is too short to give them their due. The organisms we focus on are highly diverse and relatively familiar, and they have been the subjects of our own research. My lectures in BIO120 cover the three most traditional subsets of ecology (*physiological, population, and community ecology*), but barely touch on ecosystem ecology.

The second introductory course offered through the Department of Ecology and Evolutionary Biology is BIO220: *From Genomes to Ecosystems in a Changing World*, which is offered in the winter term and typically taken in a student’s second year. As its catch-all title suggests, that course aims for a broader sweep of topics. Ecosystem ecology receives the attention it is denied in BIO120, but BIO220 also turns its attention squarely toward *human* biology. BIO220 is where you will actually consider the answers to questions like “why do we senesce and die?” It also considers a selection of human-relevant topics such as social behaviour, evolutionary medicine, the dynamics of disease, climate change, and the conservation of biodiversity.

Ultimate and proximate questions

Questions about biology fall into two major categories: *how* do the characteristics of organisms achieve life's functions? And *why* do organisms have the characteristics they do? The *how* questions are infinitely varied: in EEB courses, we tend to focus more on the functioning of whole organisms, whereas CSB courses more often look at smaller organizational levels such as organs, tissues, cells, organelles, and molecules. The *why* questions are also diverse, but have a common feature: they ultimately reduce to evolutionary questions. In this regard, it's commonplace to draw a distinction between *proximate* and *ultimate* explanations. To give an example, the wildflower *Penstemon strictus* has blue flowers (**Figure 1a**, page 1) but its close relative *P. barbatus* has red flowers (**Figure 1b**, page 1)—why? A proximate explanation would be that the biochemical pathways for the synthesis of anthocyanin pigments differ between the two species. *Penstemon strictus* makes an enzyme that sends a precursor compound into biosynthetic pathways that produce the blue and purple pigments *cyanidin* and *delphinidin*. In *P. barbatus*, that enzyme is lost; the precursor instead goes into a pathway that makes a red anthocyanin, *pelargonidin*.

This is a tidy *reductionist* explanation that tells a clear, unambiguous story at the level of genes and molecules. But if we probe deeper to seek the *underlying* cause, we can stack up further *why* questions: Why is one pathway active in one species and not the other? Eventually, we are drawn to an *ultimate* explanation that is formally very different: in the past, the evolutionary lineage leading to *Penstemon barbatus* had blue flowers that were adapted for pollination by bees (similar to those of *P. strictus*), but encountered ecological conditions in which hummingbirds provided better pollination service. In the hummingbird-rich environment, natural selection favoured red-flowered mutant plants because the red-flowered plants were more attractive to the hummingbirds. The genes that make red flowers spread through populations, replacing the blue genes. Note that the ultimate explanation is not as tidy; it is basically a complex hypothesis that rests on a series of assumptions. And most of those propositions are non-trivial to test because the events happened in the past. For some scientists, this seems untidy and unsettling. Others are drawn to the challenge. The mandates of evolutionary ecology are to pose ultimate explanations (the fun part!) and to grind through the demanding intellectual process of evaluating their plausibility (the work part!). Unpacking and explaining this process is the special province of the EEB courses. We believe that all biologists need to think about both *how* and *why*.

The broader goal In this course, we are certainly presenting you with *subject matter* that we want you to master, but we also hope that you learn deeper lessons about how science differs from rhetoric—how we get nearer to the truth by building explanations based on *evidence*. The clearest example is probably Coyne’s book, which is a masterful exposition of how one marshalls facts to decide which of two competing ideas is more worthy. We don’t have time in lectures to dwell on this aspect of the scientific method, but I hope you will see that it is a subtext that underlies all of my material. I will frequently describe some phenomenon, offer a couple of explanations for it, and then proceed to explain how some critical observation or experiment turned the tide.

As I write this in September 2017, though, we seem to have entered a new world in which former notions of evidence are in tatters. More and more people seem to be getting their information and opinions from the Twitterverse instead of from traditional sources that had built up authoritative reputations. Authority itself is now suspect. Any report that disagrees with someone’s prior conception is rejected as “fake news,” and any unwelcome scientific finding is disparaged as “junk science.” Evidence can’t penetrate very far into minds closed by the echo chambers of social media.

We are counting on you to be better than this, to be more discriminating. At this point, it is not clear what universities should be doing to counteract the tsunami of fact-free opinion that threatens to wash away civil discourse. The University of Washington is pioneering a new course on “calling bullshit” this fall; I suggest that you check out its syllabus at:

<http://callingbullshit.org/syllabus.html>

It looks to me as if you can basically “take” their course on YouTube if you are interested.

There is discussion at U of T of devising similar curricula. It’s too early to say whether such courses will succeed. But we do know that the scientific method was invented, and has evolved, as a way of forcibly checking biases against verifiable, repeatable observations. It is a strong tool if used as designed. We will try to convey that strength to you in BIO120.

Why this book?

BIO120H emerged in Fall 2010 from the breakup of its predecessor BIO150Y, a year-long course. The reorganization produced a new model in which I teach the basic ideas of ecology and Prof. Spencer Barrett introduces evolutionary biology. BIO150 had never used a textbook, relying instead on a collection of articles from the scientific and semi-popular literature. Those readings were suboptimally matched to the course's themes, and not very popular with students. When BIO120 started up, Prof. Barrett successfully adopted Jerry Coyne's book *Why Evolution is True*, but I could not find a suitable book to provide continuity and background for my lectures. Commercial ecology texts were too big and too comprehensive; students had a hard time knowing what they were responsible for. For 2014, I decided to write my own text. It seemed to be an improvement, and you are now looking at the fourth edition. I have tried to keep it short; the idea is that you will be *responsible for everything in it*, rather than having to fret about "Is this going to be on the test?" Please understand that I view this document as a companion and a complement to my lectures, not a replacement for them. You need to master *both* packages of material. Basically, the lectures are my attempt to make some basic topics *interesting*; this text is my attempt to make those topics (and a bunch of others) *clear*. Also note that the content is front-loaded; as the syllabus indicates, you need to read a lot of the book in the first couple of weeks of my lectures. The reading requirements taper off as my section progresses.

Classically, a textbook with any aspirations to scholarly excellence would be studded with references to the literature. This one is not. First, I wanted to strike a more conversational tone; second, I felt that inserting references would be a time-consuming, anachronistic exercise in futility, given that the contents of those papers are understood to "not be on the test." I expect that most of you will simply search on the Internet if you want extra background on a topic. For the few of you who might actually be seeking out key papers, you can go to full-scale ecology textbooks, or exercise the searching talents you will be learning for your writing assignment. One online resource that you may find useful is a set of modules on ecology and evolutionary biology published as *Nature Education Knowledge* (NEK):

<http://www.nature.com/scitable/knowledge/>

For the 2014 version of this book, I referred to some of those modules as required supplemental readings. Many of those articles are good, but the collection overall is uneven, poorly edited, and not integrated. Starting with 2015, I enlarged SFE so that it can stand alone.

I have also kept pictures to a minimum, given that you have access to the dozens of pictures in my lecture slides. By not duplicating that material, I can keep the file size manageable and save you ink if you opt to print this text. Therefore, I frequently refer to photos and drawings from the lecture slides, and other resources from the BIO120 course site on Blackboard/Portal. I envision you typically reading this text online, with Portal open for easy cross-referencing to pictures from the lectures. If you want a hard copy of SFE, black-and-white versions are available at modest cost. Alternatively, you may print your own hardcopy of the pdf if you like. Please use double-sided printing if possible.

An optional aid is a smartphone app (ClassApp:Biology) developed by Prof. Helen Rodd and Stephanie Halldorson specifically to help new U of T students make the transition to university biology courses. It is available from both Google Play and the iTunes store. To see what the app contains, check out: <http://classapptools.com/>

The title? Like the early ecologist G. F. Gause, I adopted Darwin's memorable phrase "the struggle for existence" for my book title because it highlights why ecology demands attention right at the beginning of introductory biology at the university level. The environment is a challenging and generally hostile place for fragile organisms. We come into being; we struggle to exist; we make a life for a while. We all eventually lose that struggle, either to accident or deterioration, although our DNA may keep going. *The struggle* has made us what we are. To understand our nature (and the nature of other organisms), we need to understand this struggle in a sophisticated way.

Secondarily, the title seems apt because many new university students find their first semester a struggle, in various senses. It should never be a struggle for *existence*, but it is likely to be a struggle with time management, with setting priorities, with finding your social niche, and possibly with grimmer issues such as loneliness, drug dependence, or depression. At the same time, you are faced with substantial challenges for academic performance and ethical standards. It can be a lot to handle. Through the colleges, the university offers a range of resources to help you through the personal issues. Do use them! On the academic side, we don't apologize for presenting you with challenges, but we do acknowledge our responsibility to make those challenges clear. Like other study aids, this little text aims to serve the goal of *clarification*.

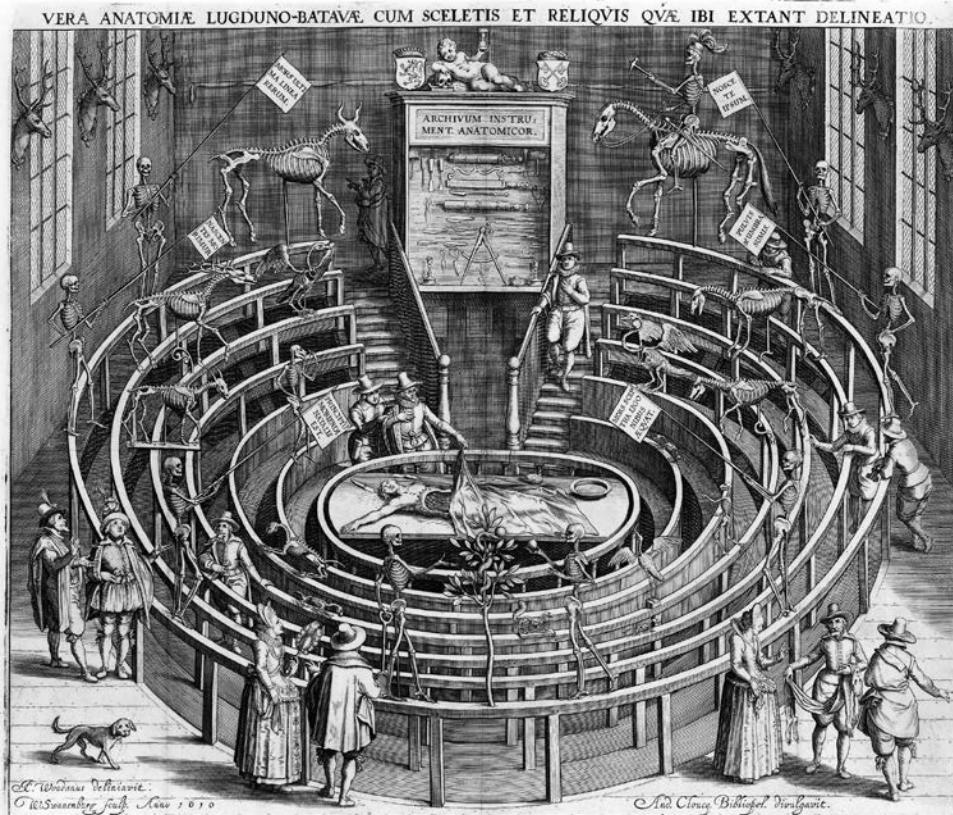
A note on pre-lecture material

When I joined the lecturing staff for introductory biology, it was a tradition to play music in Con Hall as students were arriving in the ten minutes preceding the morning lecture. I adopted that tradition, choosing to program up-tempo blues music that influenced me when I was a first-year university student on the South Side of Chicago. Some students asked for playlist details, so I started assembling artist information in PowerPoint presentations. Over the years, I have added more stuff to those loops, especially cartoons with some relationship to course material. Because I feel that U of T science students get too little exposure to the humanities, I started throwing in some poetry. Most recently, the course's focus on "sex and death" has led me to add a daily *memento mori* or "reminder of death." This tradition goes back to medieval philosophers, who supposedly had a habit of keeping human skulls on their desks as reminders of the tragic brevity of life. I think that such reminders are equally appropriate for biologists! In some extreme cases, the old-time lecturers placed skeletons among the students in lecture halls. I can't do that in Con Hall without scaring the physicists who come in at 11:00, but I will try to close each of my pre-lecture loops with an appropriate *memento mori* image. Ideally, the music will wake you up and the slide shows will give you something to think about. Admittedly, this stuff is probably more amusing to me than it is to you. But you can relax: nothing in the pre-lecture material will be on the tests!

Unfortunately, I can't play this material before the Wednesday evening lectures. There is usually a previous class that is slow to leave the Earth Sciences auditorium, and the sound system is terrible.

The good old days of biological lecturing: the anatomical theatre at the University of Leiden, 1600s, as a *memento mori*. Contemporary engraving by Willem Swanenburgh. Drawing by Jan van't Woudt

Source: Wikimedia Commons



Part 2. Adaptation to meet ecological challenges

Until 2015, my lectures on ecology preceded Prof. Barrett's lectures on evolution. Because my material depended heavily on the process of adaptation, I wrote the following section as a brief introduction to evolutionary thinking. Because we have now switched the order, some of what follows will already be familiar. Therefore, the following ideas will serve more as a review than an introduction. I have left it in, however. Concepts this important are worthy of review, and my take differs somewhat because I emphasize how adaptation is limited by constraints and tradeoffs.

We need some language to get started: the logic and terminology of function, adaptation, constraints, and tradeoffs

A founder of modern ecology, G. E. Hutchinson, published a famous 1965 essay titled *The Ecological Theater and the Evolutionary Play*. That evocative title sums up for me the way in which these two research fields are inextricably intertwined—and why BIO120 focuses on both. (I would add as an aside that the actors in that play are organisms, and that the play is never suitable for “family viewing.” It’s more like some Greek tragedy straight out of the House of Atreus. Most of the plot twists are raw matters of sex and death.)

Because ecology and evolution are so intimately tangled, you need a little bit of evolutionary background for my lectures. My lectures do not depend on a sophisticated knowledge of evolutionary processes, but the basic idea of *adaptation through natural selection* underlies everything I talk about. We need special terminology to talk about these aspects of biology. For those of you who have not yet been exposed to this way of thinking, here is my idiosyncratic primer. (This might also help with your BIO120 writing assignment.) My approach here borrows heavily from the thought and writing of my late colleague George Williams, celebrated as one of the most careful and profound evolutionary scholars of the twentieth century. George’s analysis and defense of the “adaptationist program” is most accessibly laid out in *The Pony Fish’s Glow: and Other Clues to Plan and Purpose in Nature* (1997, BasicBooks (HarperCollins)).

Williams starts his book by contrasting the pencil and the human ear, noting that the primary *purpose* of a pencil is to serve as a writing implement and that the primary *purpose* of the ear is hearing. And he argues that both of these entities became better adapted to their purpose over time through the action of *selective* processes that perpetuated better variants and eliminated variants that served their purpose less effectively. For the pencil, the process of improvement involved, first, prior *planning* by manufacturers as they devised successive refinements (for example, replacing metallic lead cores with graphite, adding integral erasers, or automating production to reduce costs), and, second, *selection*

by buyers who made purchases based on features and price. Faulty or overpriced pencils would disappear from the market. For the ear, the selection process is analogous, but only partly so. The *purpose* is there, but the *plan* is absent:

"A modern biologist recognizes no element of prior plan in the origin and evolution of the human ear. Ears and other features of living organisms...are perfected entirely by the trial-and-error process of natural selection proposed by Charles Darwin in 1859. Ears are maintained and improved because individuals with better ears are more likely to survive and pass their genes on to future generations. This conclusion is supported by evidence that organisms can have sophisticated adaptations and at the same time show...features that would not be there if intelligent planning had played a role." (Williams, 1997, p. 2).

Williams goes on to stress the message of his last sentence, by showing many examples of seemingly stupid organismal features—such as the gap between the human ovary and the Fallopian tube—that make no sense unless they represent holdovers from earlier conditions. Darwin himself realized that such characters were important in clinching the logical argument for natural selection, as Jerry Coyne further explains in the other required reading for BIO120, *Why Evolution is True* (2009, Penguin Books; review the section “Bad Design,” pp. 81-85). This is a critical point: contrary to some popular ideas, evolution should not be expected to produce *perfection*. All it does is select for new variants that perform a little better than the “previous model” in contemporary circumstances. Over millions of generations, the accumulation of those little improvements can be sublime, but telltale imperfections will always give away the unplanned, accretive nature of the process.

I will not dwell further on the evidence for natural selection, but I will be using the terminology of adaptation from here on out. Although Williams is comfortable talking about the “purpose” of an unplanned biological feature, I and many biologists prefer to substitute the more neutral word “function.” Thus, we try to understand what role a given feature (or “trait”) was selected to serve. If we can support a strong argument for how a feature increases an organism’s ability to survive and pass on its genes, we call the role a *function* and the feature an *adaptation*, and we frequently talk of such a character as being *adaptive*. So, reverting to Williams’s initial comparison, we would say that the human ear represents an *adaptation* for the *function* of hearing. It presumably arose because animals with more acute hearing had higher *fitness* than others.

In common bits of jargon, we say that well-functioning ears *confer a fitness advantage*, or that ears have *adaptive significance*.

Things can get a bit more complicated from here, though. In many animals, the shape and orientation of the pinna, or external ear, is important to acuity of hearing. Watch an alert cat swivel its ears around to track the sources of sounds, for example. But in humans, the external ear contributes very little to hearing, and certain features may have no discernible function. For example, some humans have “attached” earlobes and others have “unattached” lobes. It’s hard to imagine that one condition confers more fitness than the other, and we would tend to characterize the difference as essentially *non-adaptive*, or lacking in adaptive significance, or *neutral*. And some traits might be considered *maladaptive* because they reduce fitness. Of course, we would expect natural selection to weed out the worst of these, but remember that natural selection takes time, and it does *not* produce perfection. Maladaptive traits are likely to be prominent in a changing environment, where new ecological challenges or opportunities change the fitness consequences of pre-existing characteristics. For example, humans’ strong attraction to sweet and fatty foods was probably an adaptive trait in past environments where such energy-rich foods were rare, but in a modern, doughnut-rich environment it has become maladaptive. In addition to being lingering holdovers from different environments, apparently maladaptive traits can also be maintained by *tradeoffs* and *constraints*, which I treat below.

Also, a feature can have more than one function. Although the *human* ear doesn’t seem to enhance fitness very much through modes other than hearing, I will argue that the ears of numerous other mammals serve to regulate body temperature in addition to providing hearing. I’ll show examples, including rabbits and elephants, where well-vascularized ears provide an important way to dump heat from the body core to the environment through the flow of blood. *Nothing in the action of natural selection prevents traits from multitasking.*

Thinking about tradeoffs and constraints

Critically, natural selection does not act directly upon traits or genes in isolation. *Individual organisms* are the entities that succeed or fail, so the adaptive evolution of a particular trait will depend on its *overall* effect on fitness, through all of its functions. Suppose that a large-eared jackrabbit species (see lecture slides for photos) encounters a cooling climate in which smaller ears would allow it to retain heat better. From the standpoint of thermoregulation, smaller ears would be advantageous. However, smaller ears also might reduce hearing acuity and make the rabbit

more vulnerable to predators. That would tend to confer a fitness *disadvantage*. We would call this a *tradeoff*: because of the way this animal's ears are built, there is a negative relationship between hearing well and conserving heat well. Such fitness tradeoffs can hamper adaptive fine-tuning. Given the basic developmental blueprint for a rabbit's ear, it is probably impossible to evolve one that maximizes hearing and maximizes thermoregulation. In such cases, we will probably see structures that represent a *compromise* between functions.

Tradeoffs permeate the construction and function of organisms. Some of the most famous affect what are called *life-history* traits. Such traits concern the timing of life events, such as maturation and reproduction, and the amounts of resources that individuals invest in such functions. Discussions of life-history tradeoffs tend to invoke the economic *principle of allocation*, which simply says that resources invested in one function are unavailable to invest in other functions. This leads to propositions such as *reproduction-survival* tradeoffs (investing resources in offspring means they can't be used in maintaining the body) or *size-number* tradeoffs (if you make more seeds, they have to be smaller).

Resource-allocation tradeoffs are one form of *evolutionary constraint*, in that they reduce the range or kind of adaptation that one might otherwise see. Plenty of other factors can constrain adaptive responses, too. To return to the rabbit example, most rabbits run around in sunny, open habitats. They have famously big ears, and the species of rabbits that live in the hottest environments have the largest of all. The strong relationship between ear size and environmental temperature suggests that evolution in ear size has been largely unconstrained by other factors. But one particular rabbit, the American pika (*Ochotona princeps*; see lecture slides for photo), inhabits stony talus slopes in high mountains. It spends most of its time squeezing through crevices deep under rock piles. Pikas have very small ears. This is probably partly due to the mechanical *constraint* that large ears are delicate structures that would get worn and torn by life in the pika's claustrophobic world. Now, the mountain-tops inhabited by pikas are also cold places, so they do not need large ears to shed excess heat. However, as the global climate warms, pikas are encountering more heat stress, and the time may come when larger ears would be advantageous. But if that happens, natural selection for larger ears might be countered by the constraint posed by the wear and tear of the pika's rocky underground life style.

The basic approach of examining how different organisms meet environmental challenges in different ways is called *the comparative method*.

The comparative method is strongest when we know something about how the species are related to each other. Reverting to the *Penstemon* example, if we see two flower species that share many characteristics that suit them for pollination by hummingbirds, there are two possible explanations for the similarities. More simply, the two species could be close relatives. In that case, they are similar because they inherited the characters by descent from a recent common ancestor. In this case, we cannot count the two bird-adapted species as representing two separate evolutionary transitions from bee to bird pollination. Alternatively, the species could be very distantly related, and share so many characteristics because they have evolutionarily *converged* under selection exerted by hummingbirds. In this case, we can infer that both of them had different ancestors that were both bee-adapted, and therefore each one represents a separate, independent transition to bird pollination. In recent years, scientists have developed techniques to reconstruct evolutionary relationships (*phylogenetic trees*) from DNA sequences, thereby greatly enhancing the comparative method. All of this is treated in the evolution half of BIO120.

(Parenthetically, some of the most important algorithms and programs for analyzing phylogenetic trees have been developed by David and Wayne Maddison, former U of T undergrads who took BIO120's predecessor course in the late 1970's.)

Species' ranges as a foundational component of ecology

Part 3. Distribution and abundance as things we need to explain

Now we are finally getting to my primary subject matter. The science of ecology has been defined in various ways, and at this point in its sprawling development, it has spilled out beyond its traditional conceptual borders. However, from the standpoint of an introductory biology course, I will offer a very basic traditional definition: the study of the factors determining the *distribution* and *abundance* of organisms. That sounds very simple indeed, but construed broadly enough, it encompasses almost everything we need to consider.

Distribution is a broad term, of course, and we will use it in a few different senses. The most basic and coarsest formulation of a species' distribution is its *geographic range*, which simply comprises the areas of the planet where the species may be found. Geographic ranges are usually depicted as *range maps* (see lecture slides and **Figure 2**, page 1). To map a range, one needs to plot the positions of many sightings or collections of specimens. Such information is the special province of natural history museums, such as the Royal Ontario Museum. As such museums see cuts in funding and staff positions, we are starting to see the emergence of web-based "citizen scientist" networks as a way of collecting information about where species may be found. For an example, this site has been spearheaded by another former U of T student, Sheila Colla, to try to assess declines in pollinating insects:

<http://www.bumblebeewatch.org/>

As a general rule, ranges are thought of as rather stable parts of a species' biology. For example, they are routinely included in bird books as a guide to identification, along with details on morphology, colour, and song. For the subset of birds in our area that undergo seasonal migrations, the maps typically use different colours to distinguish the southern over-wintering range and the northern summer breeding range, which can be hundreds or thousands of kilometres apart. Given that these animals have the ability to fly so far, what keeps their ranges restricted and predictable? Why don't they just spread themselves over the whole world and appear everywhere?

There are actually a number of range-limiting factors that come into play at different spatial scales. They are partially hierarchical. The broadest and most important is *climate* and its interplay with physiology. As I will develop further when we discuss physiological ecology (Part 5), organisms will grow and survive best in places with certain combinations

of temperature and precipitation. Because of the ubiquity of tradeoffs (Part 2), it is hard for organisms to be equally well adapted to different environments. In general, an animal that is well adapted to life in hot, parching deserts will be poorly adapted to the cold, wet muskeg habitats of the subarctic. The physiological mechanisms and compensating behaviours that work best in one place simply don't work in other places. Therefore, global variation in climate is likely to draw the outermost lines that will contain a species' range.

Biomes

As we will see below, spatial variation in climate affects so many organisms in parallel ways that we see very different *sets* of characteristic organisms—termed *biomes*—in places with different climate. Different biomes tend to develop different types of soil, which greatly limit what plant species can thrive there. Very often, species range limits correspond to the boundaries between different biomes; in fact, it is precisely this tendency that allows biomes to be accurately defined by their possession of characteristic collections of resident species. Back in medieval times hardly any Europeans travelled, at least before the Crusades. Most people were unaware that other places on the Earth looked different from their homes. (When Flemish painters painted scenes from the Holy Land, the Middle East looked a lot like Flanders!) With the Ages of Exploration and Enlightenment, the western perspective broadened to appreciate this global level of biodiversity. Understanding how climate varied and thereby produced different biological communities was a significant intellectual breakthrough that we will honour by surveying some of the salient characteristics of biomes.

The most general rules of biome-level variation are simple. First, places with more precipitation develop vegetation that is taller, and usually more species-rich and productive, than places that are drier. Thus, the driest places are sparsely vegetated deserts. An increment in water allows deserts to be replaced by densely vegetated short-grass grasslands. Further increments allow short-grass communities to be replaced by tall-grass prairies, or by grassy savannahs with scattered trees. Add still more rain (and snow), and open savannahs are replaced by true forests with tall trees forming closed canopies of foliage.

Second, warmer places support bigger and more complex vegetation than colder places. As you go up in latitude toward cold polar regions, or up in elevation toward cold mountain tops, plants diminish in stature until you reach the extreme of the tundra biomes.

Third, the *seasonality* of temperature and precipitation is also important. Two regions may have the same annual cycle of temperatures and the same total annual rainfall, but if one area gets summer rains and the other gets winter rains, they will support very different sets of organisms.

To amplify that word “study,” I mean that you should learn and remember both the main *biological characteristics* of each biome, the *climatic characteristics* (annual patterns of temperature and precipitation), and how the latter influence the former. Together, **Figure 3** (page 2) and **Figure 4** (page 3) summarize the most critical knowledge. They are integrated by sharing a common colour code. Study them well, cross-refer between them, and look back to them as you read the sections about the different biomes.

Selected aspects of soil ecology in relation to biomes

Getting the map of biomes into your head is essential for developing a global perspective on biological variation. Grasping the Whittaker diagram of precipitation versus temperature lets you understand what the driving factors are. You should study this material by moving back and forth between the text and the map, asking yourself questions about *why the biomes are distributed as they are* over the continents. Some important answers should present themselves in my upcoming descriptions of global climate patterns and of physiological ecology.

In addition to climate variables, *soil characteristics* contribute strongly to differences among biomes. I have told you that precipitation affects the distributions of plants and animals, but the type of soil is critical to (1) whether that water is available to organisms, and (2) what mineral nutrients it contains. Soil influences vegetation, and vegetation influences soil. Soil is complicated stuff, and soils science is an elaborate discipline of its own, with a daunting terminology that I will mostly sidestep. Here is a very short primer.

Soil is formed by the action of living organisms and geophysical processes on some mineral substrate that we call “parent material.” Parent material is basically a geological subject rather than ecological. Parent material is often bedrock, but might also be something like sand deposited in an area by wind or water. Parent material becomes greatly altered, and added to, by (1) biological processes and (2) chemical actions such as dissolution and precipitation. Soil has a *mineral* component, some of which comes from the parent material, some of which is imported, and some of which is altered. It also has an *organic* component

that includes the decomposition and waste products of plants, animals, fungi, and microbes.

The main variation among soil types is analogous to a primary axis in human societies: *rich versus poor*. By “rich,” we mean well-suited to supporting plant growth; “fertile” is a synonym. Fertility includes three main aspects. First, fertile soils offer generous concentrations of dissolved ions of elements that plants need for growth. Nitrogen (N), phosphorus (P), and potassium (K) are needed in quantity, and are the most important *nutrients* that plants collect through their roots; there are also some necessary *trace elements*. Second, good soils also have low levels of harmful substances such as toxic metal ions (e.g., aluminium (Al), and lead (Pb)). Third, good soils offer intermediate water availability, which depends heavily on the texture of the soil particles. Very dry soils are called *xeric*, water-saturated ones are called *hydric*, and intermediate ones are *mesic*. Growing in *hydric* or *xeric* conditions requires special adaptations: water-logged roots suffer from lack of oxygen, and water-starved plants can’t conduct photosynthesis or cool their tissues (as will be discussed in Part 5).

Thought question: Interestingly, almost all plants can grow in mesic soils. Indeed, some extreme xerophytes such as cacti will actually grow better (as potted plants) in moist soils than in the dry soils where they grow in nature. Why, then, are cacti mostly absent from moist soils in nature?

Soil development: gains and losses

Soils are complex mixtures of the inorganic breakdown products of whatever rock they are derived from, plus critical additions of organic matter from animal waste and decaying dead organisms. Plant roots and leaves are usually the biggest sources of organic matter. When organic matter becomes so decomposed that its source is no longer recognizable, it becomes *humus*. The inorganic or mineral component ranges from large particles (sand) to tiny ones (clay). In some cases, wind-blown dust (*loess*) or water-deposited sediments (*alluvium*) are important. (Agriculture in ancient Egypt depended on enrichment of desert soils by annual flooding of the Nile River.) Particles of these soil components hold ions on their surfaces, and those ions are acted upon by rainwater or snowmelt percolating through the soil. Depending on the temperature and pH of the water, different ions are either dissolved or precipitated out. If rainfall is heavy, dissolved nutrients are carried far down to deep aquifers in the soil, below the reach of plant roots. This is called *leaching*, and it produces nutrient-poor soils.

If rainfall is not as heavy, water may penetrate the soil to moderate depths after a storm, but then be pulled back up by plant roots. The water is returned to the atmosphere by *evapotranspiration*. The ions stay in the rooting zone of the plants, rather than being leached downward out of reach.

The amount of leaching depends not only on the amount of water that moves through the soil annually, but also on the age of the soil. For example, it has been millions of years since the continent of Australia has experienced the kinds of geological activity that cause new rock to be exposed. Consequently, Australian soils tend to be ancient and heavily leached, even though rainfall is sparse in most places. For an example of “young” soils, Ontario can serve fairly well. The retreat of the last glaciers (something like 12,000 years ago) redistributed and exposed fresh parent materials. Tectonic uplift and volcanic activity are other important sources of fresh parent material.

Leaching also depends on the *water-retaining capacity* of the soil. Sandy soils don’t hold enough water. The spaces between the coarse silica crystals are too large to exert enough capillary force; the euphemistic description for such soils is “well-drained.” Clay soils hold water too tenaciously, leading to waterlogging that can prevent plant roots from getting enough oxygen. The best soils are *loams* that contain some sand, some clay, and plenty of organic matter. The organic content helps them to stay mesic and keeps nutrients available. Oxidized organic components make soil dark, so colour serves as a rough indicator of agricultural quality. Evocative place names like Black Earth, Wisconsin show how much the farming settlers appreciated the rich, dark, loamy soils of the Midwestern prairies.

Because temperature and precipitation are key drivers, soil development tends to vary with climate. Some generalizations: soils in cold northern regions, especially in conifer forests, tend to be tan, sandy, acid, and nitrogen-depleted (called *podsolized* soils). Lowland tropical soils tend to be ancient, red, clay-rich, heavy in iron (Fe) and aluminium (Al), and leached of good nutrients. (Iron oxide—rust—contributes the red colour.) These *lateritic* soils can support impressive forests, but tend to bake into brick after logging. They are very poor for sustained farming, and are often converted to cattle grazing because they can’t support nutrient-demanding crop plants. Podsolization and laterization both produce poor soils, but in different ways. (In mountainous areas of recent uplift, younger tropical soils derived from recent volcanic activity are less leached than laterite, and are better for crops.)

The brown soils in the temperate forested regions of Europe, China, and North America are also somewhat leached, but loamier and better suited for agriculture. And the best agricultural soils of all are the deep, black, organic-rich mesic loams developed in grasslands of central North America and the Ukraine, where rainfall is just enough to allow plant growth but not so heavy as to leach nutrients out of the soil. The perennial grasses and herbs that populated the prairies developed massive root systems that struck deeply in their quest for ground water. As those roots died and decomposed, they contributed rich humus to the soil—and not just to the surface, as leaf litter would do. Prairie soils can contain almost twice as much organic matter as forest soils nearby. Except for a few small reserves, the native tall-grass prairies of middle North America have been entirely converted to agriculture. Deep soils that have accumulated over thousands of years are now experiencing net loss because of the erosion caused by tillage agriculture.

The temperate deciduous forest biome

We will start with this biome, and then use it as a baseline for making comparisons with other types. As you read the following sections, refer to **Figures 3 and 4** (pages 2 and 3) to locate where in the world each biome occurs.

For many of us, the broadleafed trees of Southern Ontario characterize an environment whose dramatic seasonal rhythms feel familiar. Summer and winter, baseball and hockey.

How familiar? Well, we offer BIO120 in the *fall semester*. We usually use that term “fall” without a second thought, as an arbitrary term designating a particular section of the calendar—simply a synonym for autumn. If we do give it a second thought, we recognize that it refers to leaf fall, marking the transitional period between summer and winter. The typical vegetation of this biome is dominated by tall trees with thin, fragile leaves that last for a single summer before being actively shed. The most important genera are oaks, maples, hickories, ashes, beeches, etc. To anyone who grew up in the temperate deciduous forest biome (TDFB), it probably seems normal that almost all of the native plants drop their leaves at this time of year. That includes the majority of the students at U of T, because the TDFB includes the eastern half of North America, most of Europe, and much of China and Japan. We all make a great number of subtle adjustments to the changing climate, but they are so interwoven into our cultures that they don’t seem unusual. Perhaps they ought to make more of an impression! People who come from areas where the plants are green all year (“evergreen”) must get a strange feeling when they encounter their first temperate autumn.

After all, leaves are the organs that plants use to harvest energy from the environment and to conduct gas exchange. When the leaves of a deciduous plant fall off at the onset of a harsh season, those fundamental activities effectively stop. For humans, an equivalent change would be to sew our mouths and noses shut for six months every year. No eating, no breathing! That would be an extremely radical change—obviously one that would require a complete redesign of our physiology. We understand that natural selection has given plants the characteristics that help them solve the challenges of the environment. What is it about the environment in those three parts of the world that calls forth a response as drastic as the deciduous habit?

We can already guess that the answer can't be completely simple. Latitude is part of it, but not all of it. For one thing, the latitudinal belt that contains the temperate deciduous forest—roughly from 35 to 50 degrees north—also supports other biomes in some local places. Also, we find evergreen forests both below and above the latitude of the deciduous forests. For TDFB to develop, here's what's needed. There has to be a distinct alternation between warm summers and cold winters, but precipitation has to be ample all year-round. Forests need water, and more water tends to make taller forests with bigger trees, and more leaves intercepting light before it reaches the ground. Rain and snow are plentiful enough that the soil does not absorb all the water that falls, so flowing streams are common. These areas stay wet enough that forest fires are rare. The annual leaf fall contributes regular infusions of organic material to the soil. As environments get drier, however, the deciduous trees that characterize the TDFB give way to more open vegetation. The tree canopy becomes shorter and casts less shade; there are more openings or gaps where light penetrates. Eventually, the gaps dominate, resulting in vegetation in which scattered trees stand separated in sun-drenched grassy meadows. One name for such vegetation is “savannah”; some ecologists reserve that term for tropical environments, but others use it for temperate regions, also. As conditions get drier still, the trees drop out and you are in true grassland or prairie.

In North America, this happens as you move west from the centre of the forested Midwest (southern Ontario, southern Michigan, Ohio, Indiana) into the rain shadow of the Rockies. The effect of this east-west moisture gradient is striking, although the massive conversion of native vegetation to agriculture means that you can only see the transition clearly by looking at the small nature reserves that are scattered sparsely

in this part of North America. If you go north, the TDFB grades into the boreal forest biome through a broad transition zone in which broadleaved deciduous trees are replaced by needle-leaved evergreen conifers, particularly spruce and fir trees. If you go south, the deciduous forest also tends to be replaced by needle-leaved evergreen conifers, especially pines, and some other evergreen broadleaved trees such as "live oaks," so named because they do not drop their leaves as the northern oaks do.

What factors are the keys to these geographic patterns? One is snow. The deciduous broadleaved trees are best suited for environments with cold, snowy winters but long, mild summers. In those long summers, the large surface area makes large leaves excellent at capturing photons for photosynthesizing. Because there is enough rain to maintain soil moisture, water is available to prevent overheating of the leaves by supplying evaporative cooling. In these conditions, big floppy leaves are ideal. If winters were not snowy, the trees could keep their leaves all year. But winter conditions inflict terrible damage on broadleaved trees. The thin, frail leaves are torn up by frost, hail, and ice storms. Worse, whole tree branches are snapped off by the weight of snow that builds up on the large leaf area. It is better for the tree to sacrifice the delicate leaves of summer, rather than trying to retain them over the winter.

Indeed, the TDFB can be thought of as occupying a climatic "sweet spot" where conditions permit the luxury of building a new set of large, flimsy, disposable leaves each season. Farther north, the winters are too long. Snow can come late in the spring and early in the autumn, doing great harm to broadleaved plants that are leafed out. If the trees protected themselves by being more conservative—leafing out late and dropping leaves early—the short growing season in the north would not allow a long enough period of growth. So evergreen conifers, with tough, frost-resistant sclerophyll leaves and flexible branches for shedding snow, are better suited for the cold northern regions.

What about the south? In warmer climates, snow isn't much of a problem, but broad, thin leaves are prone to overheating. Especially in hot and dry areas with sandy soils that don't retain soil moisture, conditions are too dry for the deciduous trees that depend so critically on evaporative cooling. Sclerophylls gain an advantage because their smaller leaves are less vulnerable to cooking in the sun. Another factor working against deciduous trees in such areas is the frequent lack of soil nutrients. In warmer climates, warm water leaches out more soil nutrients, especially from sandy, acid soils. The lack of mineral

nutrients makes it more expensive for a deciduous tree to build a new set of leaves each season. Evergreen leaves that last several years are much less costly in terms of resources such as NPK. When pine needles finally do fall off, they form an acidifying layer of leaf litter that is slow to decay. That acidity contributes to the leaching loss of nitrogen. In contrast, the flimsier leaves of deciduous trees tend to decay quickly on the damp shady floor of cooler northern forests.

Yet another reason why pines predominate in southeastern North America is fire. Hot places with dry, sandy soils are susceptible to forest fires, and many species of pines are particularly well adapted to fire, as we will discuss later in the course. When leaf litter and fallen branches don't decay quickly, they serve as flammable fuel for fires. Smaller, local pockets of pinelands also occur throughout the more northern TDFB where local conditions favour them. Classic examples are the pine barrens of central New Jersey and Long Island, NY, where meltwater from retreating Pleistocene glaciers deposited deep outwash plains of nutrient-poor sand. Those areas burn frequently. In these local pockets, it is the poor soil, not climate, that shifts the vegetation to pinelands. Taller deciduous forests on richer soils in the same regions experience the same climate. But they are much less likely to burn because they stay moist in all but the driest summers.

Another place where chronic fire is critical is the forest-prairie border. Grasslands are dry enough to burn frequently. The main prairie grasses and herbaceous plants are all capable of being burned to the ground and then resprouting quickly from underground buds. These fires kill most trees, although a few species such as bur oak can occasionally get established and persist as scattered individuals in savannah-like grasslands. Most young bur oak seedlings don't survive ground fires, but sometimes enough years elapse between fires for the young trees to develop bark that is thick enough to protect the trees' sensitive cambial tissues from fire damage. A mature bur oak has such thick bark that it is virtually invulnerable to grass fires, allowing it to persist in a savannah-like community. Other fire-resistant tree species play similar roles in other fire-prone regions, e.g., acacia trees in Africa, ponderosa pine in the Rocky Mountains, and various eucalypts in much of Australia.

The effect of these fires is to push the grassland-forest border further into areas where the local climate would favour forests in the absence of fires. In southern Wisconsin, for example, aboriginal peoples actively set fires to maintain open grassland that they preferred for mobility and hunting. When European colonists arrived and suppressed fires, the

open grasslands that they encountered quickly reverted to dense thickets of young trees. Similar things happened in Ontario, and even in Toronto. The managers of High Park are currently trying to restore the savannah character that the area had before European settlement.

Structure of temperate deciduous forests

Trees cast shade, so forest interiors are dark. Light becomes a precious resource. Most of the photosynthesis takes place high in the leafy canopy, and trees grow so tall because they have undergone selection to vie for light and to shade out competitors. Indeed, given that soil moisture and soil nutrients are usually adequate, light for photosynthesis is the principal limiting resource for plants in these communities. The tallest deciduous forests are those that have grown on the richest, moist soils, and have not recently been disturbed by fires, wind or ice storms, or human activity such as logging. (We will discuss such disturbances later in the course.) In those darkest forests, there is very little photosynthetic tissue below the upper canopy. In more open forests, however, there may be a well-developed understory of shrubs. In both types of forests, there is a special group of understory herbs with a lifestyle found only in deciduous forests: the spring ephemerals. These are small perennial plants of the forest floor. Their strategy is to emerge and produce flowers as soon as the snow melts, before the canopy trees have leafed out. By starting so early, they can harvest light for a brief period before their habitat becomes too shady. Their entire aboveground life cycle is compressed. After the tree canopy closes in, these plants typically stop flowering, shed their leaves, produce seeds as quickly as possible, and disappear. They persist as dormant, underground roots or bulbs for the rest of the year; they may be underground for ten months and active for only two. A typical example is the trillium, familiar as the provincial flower of Ontario. In an undisturbed deciduous forest, the simultaneous bloom of numerous spring ephemeral species creates a short-lived but spectacular “carpet of flowers” that is seen in no other biome type.

Animals of the deciduous forest

The sorts of animals found in different biomes are ultimately determined by (1) the availability of plant material as food, and (2) the physical structure provided by plants. As biotic communities go, the TDFB is fairly productive. The term productivity in this sense is shorthand for the more technical term *net primary productivity* or NPP, which refers to the amount of new biomass produced annually through plant growth. Biomass is usually measured in terms of energy content per unit area per time, and a typical value for a temperate deciduous forest would be around $5000 \text{ kcal m}^{-2} \text{ year}^{-1}$. Roughly two-thirds of this productivity takes the form of wood, which is not very available as food.

The rest is mostly the production of leaves, flowers, and fruits, which are much more easily available, at least to small animals that can move around easily in the tree canopy where most of this productivity resides. The canopy is a tricky environment to negotiate, however. Compare it to a grassland, where massive grazing mammals like bison and zebras can amble around, grabbing mouthfuls of leafy forage at every step, and migrating to where the forage is most plentiful. Animals of that size simply can't operate in a forest where most of the food is presented at the ends of flimsy branches. Large mammals that do persist in the TDFB, such as whitetail deer, are actually dependent on treefall gaps and forest edges, where leaves are produced within the reach of earthbound animals. They don't feed in the dark forest very much because there are so few green tissues that they can reach. Indeed, in places where deer populations are too high, you can easily discern a browse line at forest edges: all leaves are missing up to about 2 m above the ground.

There are two consequences for the TDFB. First, most of the consumption of living leaf material is by insects rather than vertebrates. Caterpillars are especially important. Second, most of the leaf material is not actually consumed while living, but instead is shed in the fall, more or less intact. Therefore, most of the plant biomass enters the food web through decomposition on the forest floor. Much of the decomposition is done by microbes, but metazoans such as earthworms and beetles play their roles.

The abundance of small invertebrate plant eaters creates opportunities for small vertebrate predators such as mice, moles, shrews, and especially birds. The TDFB harbours plenty of insectivorous birds. Most of these are migratory—for example, warblers that arrive from the tropics just in time to exploit the spring flush of caterpillars that are themselves exploiting the spring flush of tender leaves. This food source is only available while the trees have leaves. However, there are some tough insectivores (e.g., chickadees, woodpeckers) that stay through the winter, when they glean the bark of the leafless trees for dormant insects, pupae, and eggs.

Other important plant foods are fruits and seeds, both of which are nutritionally superior to leaves. Fleshy fruits such as cherries, raspberries, blueberries, etc., are only available for brief periods of the summer. Because these sweet fruits function to attract and reward seed-dispersing birds, they typically ripen when birds are migrating. If they are not eaten, they quickly rot. Seeds, however, can be found year-round

because they preserve themselves in a dormant state. Seeds are densely packed with nutrients to support the growth of the seedlings, so they are high-quality food for animals. In particular, some of the most abundant tree species in this biome (e.g., oaks, hickories, beeches, walnuts) produce large crops of nuts that sustain many consumers through the winter.

Although birds can migrate long distances to avoid winter, that isn't an option for mammals. The TDFB harbours a few true hibernators (e.g., ground squirrels and groundhogs) with specialized physiologies that allow a profound slowdown of metabolism to save energy as they sleep underground through the winter. Some others, such as bears and raccoons, are not able to achieve true hibernation, but they do sleep most of the time in winter, and eat very little. These lifestyles depend on building up large fat reserves in the autumn; individuals that don't build up sufficient reserves may not survive the winter. Other animals, such as foxes, deer, rabbits, mice, and others, remain active all year. They typically have generalized diets, adjusting to what foods are seasonally available. Some, such as deer mice and tree squirrels (unlike ground squirrels), cache seeds and nuts when they are available, depending on recovering them later. For all of them, winter is the season of stress, when the carrying capacity of the environment is determined. Not only is food scarce, but the shorter days of winter provide less time for diurnal species to forage. Small-bodied winter residents like shrews or chickadees cannot store much fat to cushion themselves for prolonged periods. For them, each night's survival may depend on how successfully they foraged during the previous day.

The boreal forest biome

Dominated by evergreen coniferous trees instead of deciduous species, this very extensive biome is more colourfully known as the "spruce-moose" biome. In the Old World, the Russian term "taiga" is used. The boreal biome lies north of the TDFB, across Alaska, Canada, northern Europe, and Siberia. Analogous forests are hardly represented in the southern hemisphere, mostly because there is so little land mass at the appropriate latitudes. Compared to the TDFB, the soils are poor; conifer needles produce acidic leaf litter that interacts with substantial rain and heavy snow to produce heavily leached sandy soils with few nutrients and little organic matter. Coupled with cold temperatures and short summers, the poor soils render this biome unsuitable for agriculture. Therefore, human occupancy and conversion to farming is relatively low. The once-great deciduous forests now exist mostly as fragmented and isolated pockets of woodland in a matrix of cities and farmland; the boreal forest still sweeps across whole regional landscapes in the north,

except where it is frequently chopped down for logging or disturbed by fires or windstorms (as discussed in Part 8). The biome also extends southward in the mountains, where the higher elevations offer climatic conditions much like those of higher latitudes.

Although the forests are impressively dense, net primary productivity (NPP) is reduced because of the short growing season—typically four months or less—and the poor soils. Still, the evergreen nature of the trees allows them to photosynthesize on the odd warm day in early spring or late fall, so NPP can reach 2000 to 3000 kcal/m²/year.

Tree species diversity is low. A healthy deciduous forest in Ontario will likely contain 20 to 30 tree species, and more southerly deciduous forests in the U.S. may have twice as many. In contrast, large expanses of boreal forest may contain only a few: balsam fir and white spruce on uplands, tamarack and black spruce in wetlands. Numerous species of shrubs occur in wetlands and openings caused by fire, but very few in dense forest. The deep shade also prevents the recruitment of seedlings from the canopy trees. Depending on light levels, the forest floor supports a characteristic community of herbaceous angiosperms mixed in with “lower plants” such as ferns, lichens, and mosses. In North America, this low, mossy layer is evocatively named “Canada carpet.” (But the best Canada carpet I ever saw was in Sweden!)

As we will discuss under the topic of ecological succession, boreal forests are subject to large-scale disturbances, especially from fire and windstorms. As conifers grow, their lower branches become too shaded to contribute photosynthesis to the tree, and they are shed in a process called autopruning. This deadwood, plus the slow-decaying leaf litter, builds up a deep layer of fuel on the forest floor, made particularly flammable by the volatile terpenoid resins that give these trees their pungent, “piney” smell. Much of the time, this fuel is too damp to ignite, but in a particularly hot summer, it can dry out enough to become dangerous tinder. A lightning strike can ignite a massive fire. If not too much litter has built up, a burn may remain a *ground fire* that does not kill mature trees with thick bark (as with bur oaks in savannahs). If there has been a long interval without fire, though, more litter will have built up, and the fire will be hotter. If it becomes a *crown fire*, whole trees are consumed in the tremendous heat.

In wetter areas, fires are very rare, but the high water table means that the trees have shallow roots. As the trees grow taller with age, they

become unstable, and easily overthrown by wind storms. Such storms can also destroy large swaths of mature forest.

After fire or windthrow creates a large clearing, a new crop of tree seedlings will get started, all at once. These new recruits then develop into a forest of even-aged trees. Such forests have some special properties. Because so many seeds start growing at once, competition for light soon becomes intense, and all of the trees are stunted by the crowding. These conditions are poor for timber growth, and loggers scornfully refer to such stands as *dog-hair forests* that need to be manually thinned out if they are to produce merchantable trees. Even-aged stands also suffer when the trees all get old together. The previously mentioned terpene resins are the trees' principal defense against specialized insect herbivores. When a bark beetle (or a spruce budworm caterpillar) takes a bite, the pungent resin is extruded under pressure into the attacker's face. The effectiveness of this defense depends on the pressure with which the resin is squirted out, but sap pressure declines as trees age. In an even-aged stand, all the trees become vulnerable at the same age, creating a cornucopian resource for a burgeoning population of beetles. The result is yet another mechanism for a massive die-off. The mountain pine bark beetle has killed tremendous acreages of trees in the Rocky Mountain and in British Columbia. These massive kills are largely due to the low tree-species diversity of the boreal coniferous forests. Such things don't occur in the more diverse temperate forests or tropical forests, where even-aged stands are very unlikely. The damage currently being done in the West by bark beetles has been aided by recent climate warming. The longer, drier summers not only put stress on the trees; they also let the beetle populations get in an extra generation per summer.

Although boreal forests tend to be dominated by coniferous spruce and fir trees, one deciduous angiosperm tree species—*aspen*—can thrive, especially at lower elevations and latitudes. Like the conifers, aspen is pollinated by wind, and its seeds are dispersed that way, too. Unlike the conifers, aspen also shows clonal or vegetative reproduction, in which one tree can produce others at the end of lateral roots. Whole hillsides of hundreds of aspen trees can constitute a single genetic individual, arisen from a single seed. Another important group of woody angiosperms are the willows. In milder climates, these are mostly large shrubs, 2 to 3 m high. They can't grow well in dark forest understory, and their persistence in the landscape is largely due to the action of beavers. When beavers dam up streams to create large ponds, the rising water level kills established conifer trees, opening up habitats to full sunlight.

Shrubby willows are well adapted to thrive in hydric soils, so they take over from the killed trees. Because willow bark is a critical food source for beavers, these rodents are effectively creating willow farms for their own food supply. Along mountain streams, beaver families can create staircases of ponds that descend along the river beds for many kilometers. Eventually, they abandon particular ponds; the water finds its way through the deteriorating dams, and the areas dry out to allow recolonization by trees.

Animals of the boreal forest

As in deciduous forest, the summer-active insect herbivores of the boreal zone support diverse populations of migratory insectivorous birds; in the New World, the wood warblers are especially prominent. Frugivores (eaters of fleshy fruits) are less prominent than in deciduous forest, but the boreal forest offers a new feeding niche for granivorous birds like pine grosbeaks and crossbills that specialize on conifer seeds. Those seeds are available year-round, but are hard to remove from the tough, woody cones, so these birds have strong, specialized bills. Unlike the insect or fruit eaters, they don't migrate south for the winter, although they sometimes move around within the boreal zone.

Turning to mammals, the boreal forest (including its southern extensions into the mountains) harbours populations of large mammals such as deer, elk, and moose, plus their predators—wolves, mountain lions, and lynxes. We tend to think of these animals as being restricted to “the wild north,” but most of them (and the beaver) were formerly distributed through most of the temperate deciduous biome, too. Present-day patterns represent the elimination of these animals by hunting in the places where human settlements prevail. The most abundant large wild mammal in our area, the white-tailed deer, is an exception, but only barely. Whitetails were hunted so heavily by European settlers that by the early 20th-century, they were completely wiped out in the upper U.S. except for small remnant populations in Minnesota and northern Wisconsin. I don't have data for Canada, but they were probably driven to scarcity in southern Ontario while persisting further north. In the U.S., the impending disappearance of the whitetail finally drove the institution and enforcement of regulated hunting seasons. Also aided by land-use changes that were producing more and more of the forest-edge habitat that this species prefers, along with the eradication of large canid and felid predators, the deer populations came charging back with a vengeance. Today, government policies are designed to encourage hunters by keeping deer populations high. Hunting is a lucrative industry because of the sale of licenses, equipment, and accommodations. Most ecologists deplore the

artificially high deer numbers because both forests and fields are chronically overgrazed or overbrowsed. (Technically, grazing is the eating of leaves, whereas browsing is the eating of woody parts such as twigs and bark.) Especially palatable plants, such as Canadian yew, have become very rare because of deer browsing. Many ecologists have conducted experiments by putting up fences to create “deer exclosures.” In our part of the world, such experiments *always* show dramatic responses in terms of overall plant production and in plant species composition.

Spruce and fir needles are such indigestible food that almost no vertebrates can subsist on them. Therefore, the animals mentioned above (and other more specialized northern mammals such as the snowshoe hare) tend to do best in open areas rather than dense mature forest. They benefit from fires, windstorms, beaver activity, and anything else that tends to open up forest canopies and allow more plant growth at ground level. One specialized mammal that does fine in mature conifer forest is the red squirrel, a smaller congener of our comparatively thuggish Toronto grey (or black) squirrels. Red squirrels, like crossbills and pine grosbeaks (on the cover of the 2017 edition of this book), feed largely on conifer seeds, one of the few food resources in the conifer canopy that has enough food quality to support the energy demands of a small, active mammal. The other boreal canopy feeding specialist is the porcupine, which survives the winter (rather incredibly) on a diet heavy in tree bark. A porcupine can handle this low-quality food only because it is much larger and metabolically slower than squirrels. It can pack its caecum full and let microbial fermentation slowly extract nutrients. These canopy herbivores are preyed on by hawks, owls, and arboreal weasels: the pine marten specializes on red squirrels, and the fisher goes after porcupines.

The tundra biome – going beyond the trees

The cold-weather boreal-type forests give way to treeless landscapes as one moves to higher latitudes or higher elevations in the mountains, yielding to arctic tundra or alpine tundra, respectively. In both cases, the transition zone—called treeline in both cases—is usually narrow and obvious. In the arctic, the location of treeline often corresponds to where permafrost occurs, meaning that the lower layers of soil stay frozen all year, even if the top layers melt in the summer. The permanent ice makes the upper soil waterlogged during the short growing season, which hampers the development and functioning of tree roots. Arctic tundra in the summer is a mosaic of ponds, lakes, and bogs, dominated by low-growing plants that can tolerate water-saturated soils. Sphagnum peat moss, grasses and sedges are important, and soils are

nutrient-poor because the cold temperatures limit microbial action. Productivity is low.

Tundra food webs are based on grazing mice, specifically lemmings and voles. They are active all year, tunneling under the snow to feed during the winter. They are preyed on by a wide range of carnivores, including weasels, arctic foxes, and snowy owls. The mouse populations fluctuate, reaching very high densities and then crashing on a roughly 3 to 6 year cycle. The crashes have complex causes, but certainly depend in part on overgrazing: when the lemmings are at high density, they can consume half of the annual plant productivity. Such pressure greatly stresses the plants and reduces their quality as food. The pivotal role of these mice is evidenced by the fact that their predators may fail to reproduce during mouse crashes, or may migrate south in bad years. In such years, snowy owls tend to turn up in Toronto where they overwinter in the comparatively mild climate. Other grazing vertebrates include migratory geese and caribou. In some areas, the geese are so numerous that they damage the tundra by eating almost every plant part aboveground.

Alpine tundra starts at about 3500 m elevation at our latitude. It looks superficially like arctic tundra, but there are many differences. Permafrost and saturated soils are not involved; the inability of trees to grow at high altitudes is mostly due to the physical stress, especially from ice pellets being driven by prodigious winds. Near treeline, trees are pruned down to shrub size (called "krummholz") by wind action, and many of them have branches growing outward only on the leeward side of the trunk where some buds can escape damage (*flag trees*). Most of the flowering plants in the tundra are diminutive in stature, and quite a few are so-called "cushion plants" that form a dense hemispherical dome of leaves that allows wind to flow over the plant in a smooth laminar flow. Any branches that stick up too high are pruned off by the savage wind. Although the stems and leaves of these plants are miniaturized compared to their relatives at lower elevations, their flowers are often full-sized and very showy. They are typically pollinated by insects: bees and flies, with the latter becoming more important at higher elevations. During the brief blooming period of these plants, the alpine tundra is a lovely place on a mild summer day, although a nasty storm can always blow up with little warning. Hapless tourists are killed every summer because they venture out on tundra hikes in shorts and T-shirts. In winter, the alpine tundra is a forbidding place indeed; the extreme cold is made lethal by the intense wind. Large animals like mountain goats or sheep move down to sheltered sites at

lower elevations during the winter. Small ones shelter under the snow or in underground burrows.

The contrast between arctic and alpine tundra is greatest if you consider tundra on high mountains near the equator. The capsule description is that arctic tundra experiences one very long winter each year, but a tropical mountaintop experiences a short winter every night of the year and a warm summer every day. The main challenge for alpine organisms in the tropics comes from the rapid oscillation of those extremes of heat and cold. Hibernation and long-distance migration are not applicable solutions in the alpine—although some birds do move up and down mountainsides in the course of the day. Also, alpine environments expose organisms to extreme UV radiation, which is not much of an issue in the arctic.

The most characteristic mammal of northern-hemisphere alpine tundra is probably the pika, which is kind of a theme animal for BIO120. Other smallish mammals include rodents such as winter-active voles and pocket gophers, plus hibernators such as ground squirrels and marmots (which are basically alpine versions of woodchucks). Unlike the pika, all of these types are also found below treeline in subalpine meadows. Although these fossorial animals would be good prey for snakes, tundra habitats are too cold for reptiles. Therefore, the main underground predators are weasels, complemented by foxes, coyotes, and hawks aboveground.

The grasslands biome

If tundra is produced when the environment becomes too cold for trees, grasslands are produced when conditions get too dry (and fire-prone) for trees, whether in temperate or tropical environments. In North America at least, the boundary between forest and prairie grassland occurs roughly where the ratio of precipitation is just about equal to the *potential evapotranspiration* (PE). Evapotranspiration is the total amount of water that evaporates from terrestrial surfaces, OR is transpired from plant leaves. We specify *potential* evapotranspiration—rather than actual evapotranspiration—as the amount of water that could be lost to the atmosphere if that amount of water were available in the system. Therefore, PE is a measure of the drying-up power of the environment. When the ratio of rain to PE is about 1, the standing vegetation frequently gets dry enough to burn in hot, fast fires ignited by summer lightning. The above-ground plant material is consumed by fire, but the plants resprout quickly from underground buds that are insulated against the heat. In really dry environments (*i.e.*, deserts), PE exceeds water availability. But, as I describe in the section about soils, a rough

balance between PE and precipitation means that water doesn't percolate all the way through soils down into deep water tables. Therefore, nutrients are not leached out of the soils. Instead, they stay available in the rooting zone of the plants. Grassland plants tend to produce extensive and deep root systems, and those roots contribute organic matter (and more nutrients) to the soil when they are shed. Therefore, prairie soils are extraordinarily rich: deep, black, and beautiful.

The productivity of grasslands depends on rainfall. In North America, there is a gradient from tall-grass prairie in the East (near the forest boundary), to mid-grass prairie further west, to short-grass prairie in the dry rainshadow of the Rockies. Early settlers described the tall-grass prairie in awestruck terms: sometimes a man had to stand on his horse's back to see around him. The dominant species were grasses, especially big bluestem, but tall-grass prairies also contained a great diversity of forbs (broadleaved herbaceous angiosperms other than grasses), sometimes approaching 300 species in an area. Many of these were legumes that helped to restore nitrogen to the soils; most legumes harbour "nitrogen-fixing" *Rhizobium* bacteria in their roots. Those bacteria can mediate the conversion of nitrogen from atmospheric gas—which is plentiful but useless to plants—into ammonia and nitrate compounds, which plants can use. This regeneration of soil nitrogen is critical, because the nitrogen compounds in plants are converted to gaseous oxides and lost to the atmosphere when the prairie burns. (In contrast, fire converts potassium and phosphorus to mineral ash that stays in the system.)

Sadly, only tiny remnants of tall-grass prairie survive; the conversion to agriculture has been almost complete. Drier shortgrass prairies have mostly been turned into cattle grazing. Institutions such as the University of Wisconsin Arboretum have restored rich tallgrass prairies from degraded farmland, but this requires much effort. First, the native plant species have to be found and propagated. After that, maintaining grassland requires a regimen of controlled fires.

Unlike forests, intact grasslands typically support large populations of grazing mammals. This lifestyle is made possible because all of the plant biomass is close to the ground, rather than borne on high branch tips as in forests. The African savannahs are famous for their great herds of antelope, zebras, and many others, which still persist in some great preserves such as the Serengeti plains. These herds migrate, following the rains brought by the movements of the intertropical

convergence. In North America, their ecological equivalents are species like bison and pronghorns, which have barely held on in much smaller remnants. However, the fossil record shows that North America (and Eurasia) formerly had much richer grazing faunas. There were massive extinctions around the time that humans arrived, and some paleontologists believe that human hunting was responsible. Under this hypothesis, the African mammals gradually evolved and adjusted their behaviours as humans also gradually evolved and developed hunting techniques. In other places, humans suddenly arrived by migration with hunting techniques fully developed, and used that technology to slaughter most of the “unprepared” grazers.

Although the hunting hypothesis is somewhat controversial, it seems certain that humans did tend to convert forests to grasslands by deliberately starting fires as a management tool for hunting and for maintaining habitats in an open state that makes it easier to spot predators and other sources of trouble. The ecologist Gordon Orians has proposed that humans feel psychologically comfortable only in savannah habitats: when we move into forested areas, we cut down most of the trees; if we move into treeless areas, we plant scattered trees. All such speculations about evolutionary bases for human psychology are controversial and hard to test, but tantalizing. Certainly, humans' planned parks all do resemble savannahs.

The desert biome

Recall that the Big Two environmental factors are temperature and water. Tundra represents “living on the edge” with respect to temperature. Deserts occupy an analogously precarious position with respect to water. Because deserts pose extreme physiological challenges, I discuss desert plants and animals in the section on physiological ecology, so my treatment here will be brief. Basically, deserts develop when conditions become too dry to support a more-or-less continuous cover of short grasses. The plants that do persist in deserts are usually in strong competition with their neighbours for water, *not* for light, and the survivors are species with specialized adaptations to acquire water and to use it efficiently. Substantial areas of unvegetated soil surface open up, and the soil is poor stuff: sandy or gravelly with very little organic matter and little nitrogen.

Deserts may be hot (*e.g.*, the Sahara) or cold (*e.g.*, the Gobi). (Indeed, the polar regions get little precipitation, and much of Antarctica would have desert-like organisms except that it's too cold to have organisms at all.) I'll focus mostly on hot deserts because high temperatures exacerbate the

water stress that organisms must endure, provoking more dramatic adaptations.

In the very driest deserts in the world, such as the Atacama in South America, there is simply very little life. The plants are usually very small, and grow very slowly. (My candidate for the “most austere flowering plant” would be *Lithops*, a genus of cactus-like plants from Namibia that evade being eaten by looking like pebbles. Check out pictures on the Internet.) The biology gets much more exuberant in less severe deserts (like the Sonora) in which the persistent drought is relieved during one or two brief wet seasons. The wet seasons may come from rain, as the intertropical convergence passes over, or from snowmelt that runs down watercourses from nearby mountain ranges. The transitory availability of water gives the inhabitants of these deserts something precious to fight for and to hoard. Plant growth forms are much more diverse in those deserts than they are in the somewhat wetter shortgrass prairies. The prairies support mostly bunch grasses, with small aboveground shoots and deep root systems. Some of those grasses get into the deserts, too, but they are joined there by many microphyllous shrubs and succulents, treated in my lectures on physiological ecology. The many open patches of bare soil also open up opportunities for another important class of plants, desert annuals (also mentioned in lecture). These short-lived plants wait out one or more long dry seasons as dormant seeds. If the winter is wet enough, the seeds germinate and the plants rush to grow, flower, and set new seeds in a few weeks. Many of them have water-soluble germination-inhibiting compounds in their seed coats. This prevents them from germinating unless they have been exposed to enough water to dissolve away the inhibitor—an elegant mechanism for coordinating germination with especially wet years. In the early spring after a wet winter, the concentrated blooming of annuals makes the Sonoran desert look like a lush garden. By June, all of the annuals have died, and it once again looks like a parched desert. It’s instructive to compare this *desert annual strategy* with the *spring ephemeral strategy* so common in temperate deciduous forest. Recall that spring ephemerals are perennials, not annuals.

For the most part, desert animals are just heat-adapted variants of species found elsewhere. A few large grazers, such as desert bighorn sheep, can get by in mountainous areas where they can ascend to higher elevations in the summer. Small rodents, mostly nocturnal, are common, diverse, and well-studied by ecologists. They subsist largely on seeds, and serve as prey for snakes and owls.

Under human domination of the landscape, deserts are increasing in area. Typically, moderately productive shortgrass prairie is converted to less productive desert by a worrisome process called desertification. Various factors are implicated, but the most important is probably grazing by domesticated livestock, especially goats. Goats are versatile foragers that can somehow scrape up a living in very forbidding habitats, but they tend to leave little plant life behind. Despite their hooved feet, they are adept at climbing trees to strip even the highest leaves. This makes for comical photos (see the Internet), but the ecological effects are grave. Climate warming is also instrumental in driving the expansion of hot deserts, with bad implications for human societies in these unproductive but still overcrowded areas. Fire plays almost no role in deserts simply because the plants aren't productive enough to produce the critical amounts of fuel needed to sustain a burn.

Shrublands: chaparral vegetation and the Mediterranean climate

This biome has strong affinities with seasonal deserts, but is wetter and more productive. The key climatic factors are hot, rainless summers and warm, rainy winters with negligible frost. Such conditions are found in several places around the globe, including much of the Mediterranean basin, much of California, and some parts of Africa and Australia (as always, see **Figures 3 and 4**, pages 2 and 3). The most characteristic resulting vegetation has different names in each country: *chaparral* in North America, *maquis* in France, *matorral* in Chile, *fynbos* in Africa, *mulga* in Australia, and many other names for other variants. Bush or scrubland are other general names. The main characteristics that differentiate this biome type from deserts are, first, that the dominant plants are woody shrubs or small trees. Second, the wet winters allow enough plant growth that there is plenty of fuel to sustain large fires during the hot summers. Plants typically have the sorts of adaptations to fire that we will discuss later in the course; most commonly, the stems of perennial woody plants are destroyed by hot fires, but the root systems survive and quickly send up new shoots. These plants are well-adapted to fire, and the consequences of burning are rather small. Interestingly, these climatic conditions are particularly good for wine grapes!

Fires have harsher consequences for humans: despite the omnipresent fire danger, some of these areas are heavily populated, especially in California, Europe, and parts of Australia. As I write this in September 2017, California and British Columbia have again suffered through major fires, even though California finally broke out of its longest

drought in memory. Indeed, fire is so omnipresent in that chaparral state that it has its own webpage:

<http://www.fire.ca.gov/general/firemaps>

During the recent drought, five years of below-normal precipitation left much of the vegetation dead, dry, and flammable. In 2015, the snowpack in the Sierra Mountains was estimated to be at the lowest level in the last 500 years. This drastic situation is caused by several factors, including an El Niño weather pattern, hot weather probably exacerbated by anthropogenic climate warming, and short-sighted fire control policies. In the U.S., authorities respond to immense pressure to extinguish any fires near human habitation. In consequence, fuels accumulate to the point where larger, uncontrollable fires become inevitable after years of suppression. Across the border in Mexico, there is less money for firefighting, and most small fires are allowed to burn themselves out. This prevents the sort of fuel buildup that causes the catastrophic fires in the U.S. In Mexico, fires are more frequent but usually trivial in extent and damage.

The tropical rainforest biome

To a first approximation, we can understand tropical forests as being akin to their temperate counterparts but without any winter. With the stress of cold weather taken out of the equation, “the Big Two” factors reduce to one: water. Thus, the amount and the seasonal timing of precipitation are the key factors for tropical forests (at least in the lowlands: big tropical mountains present a more complicated story because cold re-enters the equation).

Much of the tropical zone does experience distinct wet and dry seasons due to the advance and recession of the intertropical convergence. Nevertheless, there are places, such as the Amazon basin and parts of western Africa, where the *intercontinental convergence zone* (ITCZ) moves very little. It just sits there making rain. Those places produce the iconic, idyllic, romantic rainforest, more technically described as *evergreen lowland forest*, where at least some rain is likely to fall every day. Rooted plants are never water-stressed, so their fitness depends on their ability to compete for light and nutrients, especially the former. The relentless competition for light pushes trees to grow taller and taller, so the canopy of a well-developed tropical rainforest is much higher than that of most temperate forests. The interior is dark and intensely shady. The first European visitors were awestruck by these “green cathedrals.” To less reverential Europeans, the majestic stature of the dominant trees also suggested that these forests could profitably be converted to very productive agricultural fields, but that inference turned out to be wrong.

The heavy rainfall produces highly leached soils, and the fact that the leaching water is *warm* means that the soils typically become enriched in aluminium and iron. The resulting soils are red like bricks, and when the trees are cut down, the clay bakes to something even more like bricks. (This is less true in forest on more recent volcanic soils.)

Compared to our “baseline” temperate deciduous forest, rainforest interiors are darker because more layers of leaves intercept the sunlight. In contrast to Hollywood images of hardy explorers wielding razor-sharp machetes to slash their way through impenetrable jungles of green hell, intact tropical forest floors are actually fairly open and easy to walk around in. Too little light filters down to allow dense growth in the understorey. The plants that do occur are specialists that are extremely shade tolerant; many of them are so-called “giant herbs” with large leaves that would quickly overheat if exposed to full sun. You have surely encountered some of these as indoor houseplants (“dentist’s office plants”!), a role to which they are preadapted. Despite the forest’s great productivity, there is little leaf litter on the floor: the warmth and the unrelenting humidity mean that fallen leaves quickly rot into humus. Now, those impenetrable green jungles do occur, but only at forest edges or in places where the canopy has been opened up by a falling tree.

Other differences: the pervasive moisture fosters a diverse community of tiny plants that grow on the leaves of larger plants. These algae, mosses, lichens, liverworts, and others are collectively called *epiphylls*. Eventually, they may completely cover the hosting leaf, shutting off photosynthesis. Rainforest leaves frequently are smooth and shiny, and come to a downward-angled point called a *drip tip*. These features are thought to retard the establishment of the epiphyll crust. Plants from many different families produce such leaves, a good example of *ecological convergence*.

Other tropical plants that are specialized for a moist and light-limited environment are *epiphytes* and *lianas*. I’ll conduct a demonstration of epiphytes in lecture: these are full-sized plants that grow with their roots attached to the trunks or branches of other plants rather than in the soil of the forest floor. They depend on getting their seeds dispersed into humus-filled crotches of high branches. Lianas are woody vines (the things that Tarzan swings on) that are rooted in the ground but can quickly climb up host trees to reach the canopy. Both of these life forms represent “cheap” ways for a plant to get its photosynthetic machinery

up into the sun of the forest canopy without the expense of building its own trunk. Both strategies depend on wet tropical conditions.

The most celebrated feature of tropical rainforest is the *elevated species diversity*. A well-studied plot in Panama, 50 hectares in extent, hosts 300 species of woody plants. And there are even richer rainforests elsewhere, for example in the Amazon. This super-high biodiversity is paralleled by other types of plants and by most groups of animals, including birds, reptiles, mammals, leaf-eating insects, and countless others. (Interestingly, my favourite group of animals—the bees—actually are most diverse in desert and Mediterranean habitats.) The general pattern of high diversity in the tropics, tapering off to low diversity at higher latitudes, has fascinated ecologists and evolutionists, and numerous hypotheses have been advanced. (You may recall that Spencer Barrett cited this phenomenon as one of the “big questions,” along with the evolution of sex, in his first lecture.) There is no consensus regarding the “correct” answer, but it seems certain that *part* of the difference is simply that core tropical areas have been undisturbed and stable over past geological periods while the temperate zones were being devastated by cycles of glacial advances and retreats. In lecture, I’ll discuss one other hypothesis that has received considerable support. This idea was published simultaneously and independently by Daniel Janzen and Joseph Connell, and bears both of their names. The *Janzen-Connell hypothesis* supposes that the lack of winter in tropical areas means that plants never gain respite from the attacks of herbivorous insects and other enemies such as fungi. Most of those attackers are specialists that can only attack one species of plant. Therefore, if a plant species starts becoming abundant in any one locality, its natural enemies will also build up and prevent it from reproducing by attacking its seedlings. Therefore, no single plant species can become dominant, and many can coexist. Quite a few studies have supported the central premise of the J-C idea by showing that tree seeds and seedlings do experience higher mortality in the vicinity of their parent. For a fuller consideration of other ideas, take more advanced EEB courses!

Animals in tropical rainforests

In general, most types of animal that live in temperate forests are also found in the tropics. Thus, rainforests host familiar types such as squirrels, porcupines, and deer. But there is another set of new tropical specialities. In tropical forests, most species of plants are adapted for *pollination and fruit dispersal by animals*, rather than wind, which is used by many temperate-zone plants. In part, this is because not much wind penetrates the interior of a tall evergreen forest. The year-round availability of fruit, especially of some special plants like figs, opens up a

new feeding niche that is absent in temperate forests. Specialized fruit-eaters (frugivores) include birds like tanagers, parrots, and toucans, and mammals such as certain bats and monkeys. Similarly, other birds and bats can specialize on feeding from flowers around the year, without having to migrate to avoid winter's dearth. The year-round superabundance of foliage also creates opportunities for some midsized arboreal folivores such as sloths, and leaf-eating primates like howler monkeys in the Americas or colobus monkeys in Africa. Unlike the more typical omnivorous monkeys, which are ceaselessly foraging for rare, high-quality foods, the leaf-eaters tend to fill their stomachs quickly and then bask in the sun while their gut microbes slowly digest their more refractory diet. Gorillas are the ultimate example of this laid-back vegetarian primate lifestyle.

The warm climate favours ectotherms, especially reptiles, amphibians, and an impressively diverse array of insects. Small arboreal lizards and frogs join birds as important insectivores that patrol the leafy canopy. Ants are everywhere, and some of them take on new roles that temperate ants don't fill. For example, *leafcutter ants* gain their food from leaves of many species of trees. Those leaves are defended by many different toxic compounds that the ants can neither tolerate nor detoxify. The ants get around the nasty chemicals by bringing leaf pieces home to humid underground chambers where they grow fungus on the leaves. The fungus has the necessary chemical arsenal to grow on the leaves, and the ants eat the fungus.

First-time visitors who have been primed by stories of the hyperdiverse floras and faunas of rainforests are frequently disappointed on encountering the real thing. The problem is that most of the action—the fantastic orchids, the flowering and fruiting trees, the colourful birds, and the chattering monkeys—are usually carrying out their business in the sun-drenched canopy, 30 to 50 m above ground. Although you, the earth-bound human, can hear some of the chatter and songs, you mostly just get frustrated as you peer into a wall of green—with sweat-filled eyes through fogged-up binoculars. The recent vogue in tropical forest ecology, if money permits, is to erect giant construction cranes like those that dot the downtown Toronto skyline. By swinging the boom around and lowering a cage on the cable, a human observer can actually get to where things are happening.

The tropical deciduous forest biome

Where tropical areas experience sharply differentiated wet and dry seasons, we often find deciduous—rather than evergreen—forest. These trees are *drought-deciduous*, not *winter-deciduous* like those in temperate

deciduous forest, but the underlying cause is analogous. It is simply less costly for a tree to drop its leaves for the dry season than it would be to retain them and have to keep them supplied with water. In these forests, you can walk around kicking your feet in a layer of fallen leaves—just as in the temperate deciduous forest but *not* as in an aseasonal rainforest. Of course, not all tree species have identical ranges of tolerance. Where dry seasons are both harsh and long, virtually all of the species may be deciduous. But in less extreme conditions, certain species may drop their leaves while their more tenacious neighbours retain theirs. Trees in *riparian* habitats—along streambanks—tend to keep their leaves.

Interestingly, many species produce their flowers during the dry season even though they have dropped their leaves. Presenting flowers while the leaves are gone has some advantages: the flowers will be more visible to pollinating animals. (And for the minority of species in these habitats that are wind-pollinated, pollen can be carried further if there are no leaves to block the breeze or to intercept pollen grains.) Furthermore, if the plant can also mature its fruits before the rains come back, having fruits exposed will also aid dispersal of the fruits by either animals or wind. True, making and presenting flowers during the dry season will be costly in terms of the water needed to expand the petals and supply the flowers with nectar. However, compared to leaves, flowers are cheap, disposable units that are never built to last. It's easy for a structure to present a good cost:benefit ratio when its cost is small.

A couple of biomes that we're basically ignoring

I've only hit the high spots. For completeness, let me mention (1) temperate rainforest and (2) tropical thorn scrub. I skipped the former because it occurs only in a few small areas around the world where one finds cool-cold winters coupled with continual rain. The wettest of these are big, spectacular forests with giant, moss-covered trees that need to be seen to be believed. If you travel to the Pacific Northwest, try to get to the Olympic Peninsula of Washington State. The Otway Peninsula in southern Australia is also spectacular. Still, these extreme areas tend to be small and localized on a global scale. In **Figure 3** (page 2), the Pacific Northeast rainforest is so small that it only has room for one bar of its stripey colour code. Cape Otway is too small to appear at all.

Unlike the rare, boutique biome of the temperate rainforest, tropical and subtropical thorn scrub is very extensive in dry places like parts of Africa, Mexico, and South America. Basically a variant of chaparral, it grades into tropical savannah/grassland and hot deserts, but gets relatively little study and little tourism because most people simply find

it unappealing, if not positively repellent. Like all biomes, it hosts plants and animals with exquisite adaptations and fascinating stories to tell, but still. “Thorn scrub”—enough said!

Concluding remarks on biomes

I regret that I don't have time to treat these topics in lecture; I would enjoy doing that! I hope that this textual account has provided some organismal biology to enrich the bare-bones proposition that *“the amounts and seasonality of temperature and precipitation determine the biomes.”* I want you to think about biome variation from the perspective of how different combinations of those factors make certain areas of the world hospitable to some organisms and hostile to others. Further, I want you to see how *phenological timing* is as important an aspect of an organism's phenotype as its morphology. Finally, I hope you are thinking about these things from the evolutionary perspective of how environmental challenges affect the success or failure of an individual organism.

A few thought questions:

1. *Where are the four major areas of Mediterranean or chapparal shrubland in the world, and what are the common features that have driven their convergence?*
2. *Why are the great conifer forests restricted to the northern hemisphere?*
3. *What are the similarities and differences of biome distribution in the two southern-hemisphere continents, Africa and South America?*
4. *To what extent are the biomes determined by latitude alone?*
5. *When you see different biomes at a single latitude (e.g., at the equator in Africa or at 40°N in North America), what more local factors are at work?*

Smaller-scale variation

Within biomes, there is further subdivision into finer units called *habitats*. For example, within the coniferous Boreal Forest biome in Canada, some localities are dominated by white spruce and others by black spruce. Other species of plants and animals are non-randomly associated with the two spruce species, creating a lower level of variation that you can think of as mini-biomes within full-scale biomes. Habitat-level variation depends on smaller-scale differences than regional climate, though: swampy lowlands versus drier uplands, acid soils versus alkaline soils, south-facing slopes versus north-facing slopes. Within a biome, these different habitat types experience similar annual rhythms of temperature and precipitation, but they still differ in smaller-scale characteristics that profoundly affect the distributions of organisms.

Within the ultimate limits imposed primarily by climate, and secondarily by habitat-within-climate, still other factors—lower in the hierarchy—come into play. Most of these are not driven by physiological limitations. Rather, they arise from animals' *behavioural choices* that are lumped under the term *habitat selection*. Many animals, for example, depend on particular plant species or plant communities, either for food or habitat. In treeless areas of the Sonoran Desert, for example (see lecture slides), the miniature Elf Owl nests in cavities that woodpeckers excavate in the large stems of saguaro giant cactus. Woodpeckers vacate their cavities after one nesting season, leaving behind a critical resource for the owls and other secondary occupants. But this means that you will see Elf Owls primarily in the subset of desert habitats that offer saguaro (or other big columnar cactus) *and* woodpeckers. The owls *choose* to avoid non-desert habitats as unsuitable. But you can tell that they do not depend on saguaros in any obligate physiological sense, because they can successfully nest in oaks or cottonwoods in those moister desert habitats that do support woody trees.

Sometimes, organisms exhibit extreme habitat selection for no apparent reason. In lecture, I will mention Kirtland's Warbler, a small insect-eating bird that nests *only* in forests of jack pine that have grown up after fires that occurred between 5 and 20 years before. This bird is a *habitat specialist*, and its hyper-special habitat is very rare, especially now that European settlement has reduced the natural frequency of fires. This bird's summer breeding range is now restricted to tiny areas in western Ontario, Michigan, and Wisconsin, and it is teetering on the brink of extinction simply because it is so absurdly picky. But is it possible that the cause is something other than behavioral preference? Is this warbler somehow constructed so that it can't survive in any other type of habitat? We can't absolutely rule out that possibility. We can, however, consider it *highly unlikely* because closely related species of warblers—which are identical in size, shape, and feeding behavior—have ranges that cover almost all of North America: they are *habitat generalists*. This suggests that Kirtland's Warbler is probably cognitively limited. Some hard-wired circuit in its little bird brain rejects as unsuitable any habitat other than young jack pine forests. Presumably, that neural circuitry had adaptive value in past environments when pine forests were common and extensive. It appears to be *maladaptive* in the transformed landscape of today.

Thought question: Suppose you were assigned to rescue Kirtland's Warbler from extinction. Obviously, creating more young stands of jack pine would help, but let's say that those options are limited because you can't afford to buy up private land. What else might help? Could you do anything to make the birds themselves less finicky? If this exercise intrigues you, you might enjoy this Radiolab segment that describes various conservation efforts, and how one man died in the effort to save the warbler:
<http://www.radiolab.org/story/91723-weighing-good-intentions/>

When ranges shift

I mentioned that ranges are usually viewed as stable characteristics of a species, but they do shift over time. For example, bird guides from the early 20th century cite the familiar Cardinal as a permanent resident of the southeastern United States, but this species has been moving northward for decades and is now a common sight year-round in Toronto. Northward shifts characterize many other species in our part of the world. Much of this movement is attributed to a warming climate with shorter, milder winters. In the case of the Cardinal—a bird that happily adopts towns and cities—an increase in bird feeders has probably been more important.

Climatically-induced range shifts by *resident* species are usually slow. In contrast, dramatic range shifts can occur rapidly when a *new species is introduced* to a region or a continent where it has previously been absent. If the conditions are suitable, such organisms may become *invasive species*. Such species typically spread so rapidly, and reach such high densities, that they become pests, frequently having harmful effects on native species. A well-documented example—sticking to birds—is the Common Starling. In 1890-1891, about 100 birds of this European species were deliberately released into Central Park, New York City, by an eccentric enthusiast who was determined to introduce to America all of the birds mentioned by Shakespeare. It took these birds a while to get established, but then they spread fast. Descendants of the original birds reached the Midwest by 1928 and California by 1942. Today they are abundant almost everywhere in North America. They are widely regarded as aggressive, messy nuisances that compete with native birds for food and nest sites. You will see them on the Sidney Smith patio, scrounging for hot dog scraps alongside another introduced urban specialist, the House Sparrow. Invasive species of plants and all sorts of animals are becoming ever greater problems as intercontinental travel moves more of them into new locations.

Thought question: It's common for New World cities to have been invaded by species native to the Old World, but rare for Old World cities to be invaded by species native to the New World. What explains the asymmetry?

It is not always clear why invasive species run so wild in new habitats. A favoured theory is that the invasive populations have left behind natural enemies such as parasites and diseases that controlled their populations in "the old country." This is a very attractive hypothesis, and it has clearly happened in some cases. It is less clear whether it constitutes a general pattern; evidence is mixed.

Habitat and the ecological niche

The literature is full of references to the "ecological niches" of different species. Unfortunately, this is one of those terms that has become so attractive that it is used in various different senses. When you read it or use it, it is essential to clearly establish the meaning that is intended. Initially, the term was used by zoologists, although it has been broadened to include plants. When Joseph Grinnell coined the term in the early twentieth century, he intended to describe mostly the *particular microhabitats* that a particular animal species frequented. To a lesser extent, he included the characteristic *behaviours* that the organism employed in exploiting its environment. The intention was to be comparative, to indicate the place in its community that a particular species occupied, relative to others.

In the 1920's, Charles Elton promulgated a concept of the niche that downplayed the physical location where an animal would be found, and emphasized almost entirely its *functional role* within the community. He considered an animal's most important functions to be its "food relations"—whom it would eat and who would eat it. He viewed niches as akin to *jobs* in human society. He wrote, memorably, "When an ecologist says, 'There goes a badger,' he should include in his thoughts some definite idea of the animal's place in the community to which it belongs, just as if he had said, 'There goes the vicar.' "

In the 1950's, G. E. Hutchinson developed a catch-all niche concept that could be expanded to both function (job) and habitat (place), and pretty much any other ways in which one could characterize differences among species. He used a geometric analogy of "ecological dimensions." He envisioned plotting *species use of resources or habitats* on linear axes. For comparing the niches of insect-eating birds, for example, one axis could be the size of insect prey. If one bird species ate a size range of mostly small insects and the other ate a different size range of larger

ones, you could plot both ranges on the “insect-size” axis and determine how much differentiation there was. The ranges of the two birds on this one axis would be components of their ecological niches. You could then erect a second axis, perpendicular to the first, that represented some other aspect of resource use, say the height of vegetation in which the two birds foraged. The birds might use different portions of that axis, too (see lecture slides). Here the first axis is consistent with Connell’s conception of the niche as a job; the second axis aligns with Grinnell’s focus on microhabitat use. A third axis could be time of day. A fourth could be the size of perches used while foraging. Dimensions could also include climatic limitations of the sort we discussed above as ranges of tolerance. By successively adding axes, we create an *n-dimensional hypervolume* of “niche space” in which each species would have a characteristic location. Human brains are poorly equipped to visualize or to diagram niches in more than three dimensions, but we have mathematical tools for describing the breadths and overlaps of hypervolumes in n-dimensional space. Species that are close to each other in niche space have similar roles in the community. Those that are far apart, or have no overlap at all, are very different *in a way that comprehensively takes in many ecological aspects*.

We will return to this idea in Part 7, when we examine a 1970’s approach to community ecology that depended heavily on trying to quantitatively measure Hutchinsonian niches. That approach has lost momentum, for illuminating reasons that we will discuss. For now, my personal inclination is to be somewhat wary of the ecological niche. It is more like those seductive, Zen-like buzzwords in psychology and the humanities that people love to debate about, partly because the discussions are impossible to settle without agreed-upon definitions. For example, ecologists have debated whether a niche can exist without a species (like a job without an incumbent), or whether a niche can exist only as a property of the species that embodies it. To a pragmatist like me, this resembles fretting about the sound of one hand clapping. Still, ecology is stuck with the notion of the niche. We just need to be extremely alert to the context in which it is being used.

Modern usage of the niche concept

Ecological niche modeling is a relatively new field—and a very active one!—that is receiving much attention in the context of global climate change. Suppose that a weedy plant has been introduced to Australia from South America, and you want to predict what range that weed will be able to occupy in Australia. One approach is to look at the temperature and precipitation levels that determine the boundaries of the species’ range in South America. Essentially, you are inferring that the spe-

cies' geographical range is (and will continue to be) determined by its ranges of tolerance. Then you map where those limiting temperatures and moisture levels occur in Australia. That map is your prediction of the likely limits to the expansion of the invading species. In the trendier area of climate change, this kind of limit mapping is used to predict how far a species might have to migrate to keep up with climate change. That is, you take a simulation model of future climate, and see where future conditions match the conditions currently found at the current limits of the species' range. Typically, the predictions will be that the species will need to migrate poleward in latitude, or upward in altitude. You can also use the simulation model to estimate how fast the climate will shift, which lets you consider whether the organism you are studying can actually migrate that fast.

Thought question: Ecological niche modeling is an elegant concept, and building the models is an exciting challenge for quantitative ecologists to collaborate with climate scientists using massive databases. But what are the implicit assumptions of the process? How might violations of those assumptions compromise the resulting predictions? Of the various conceptions of the ecological niche, which aspects are captured by this modeling process?

Part 4. Patterns of climate from global to habitat scales

From here on out, this text will adhere more closely to my lectures. In lecture, I will develop an account of atmospheric circulation patterns, using a number of very important illustrations that I do not reproduce here. Refer to those lecture slides to make sense of this brief and supplementary verbal version! If you are shaky on basic descriptions of the earth, such as latitude and longitude, you may need to review those.

Latitudinal patterns: how the sun heats the earth

The environmental variable most important to organisms is temperature. It follows that the single most important source of ecological variation for life on earth arises from the latitudinal gradient in temperature: hot near the equator, grading toward cold at the poles. This pattern arises from the uneven distribution of radiant electromagnetic energy that is continuously blasted at the earth from the sun. Note that this distribution would *not* be uneven if the earth were a cylinder spinning on its axis rather than a sphere. Why? The spherical shape means that the sun's rays strike the earth at different angles at different latitudes. The sun is far enough from the earth that we can consider the incoming photon flux as a uniform stream of parallel rays. Making that assumption, the density of photons is highest at the equatorial regions, where the surface of the earth is essentially perpendicular (90°) to the vector of the incoming photons. (Actually, the angle shifts with the seasons, but the earth's surface at the equator *is* perpendicular to the sun's rays if we average its position over the course of a year.) Photon density per unit area declines as we move toward the poles because the angle of incidence declines from 90° to 0° , eventually reaching a tangent, where the rays skim parallel to the ground at the poles, not delivering any energy to the surface but instead zinging ineffectually off into space.

Seasonal variation in climate is immensely important to organisms. It arises because the earth's axis is tilted at about 23.5° off the vertical. As it makes its annual revolution, therefore, different parts of the earth experience the sun as being directly overhead at noon. At the spring and autumn equinoxes, the sun is directly above the equator. At the northern hemisphere's summer solstice, it is directly over 23.5°N , and at the winter solstice, it is directly over 23.5°S . The annual shifts in angle are gradual, but their effects on daylength and on heat input are profound. The latitudinal lines around the earth at 23.5°N and 23.5°S are called the *Tropic of Cancer* and the *Tropic of Capricorn*, respectively. The belt bounded by them is called the *tropical region*, or just the "tropics." The "*solar equator*" (i.e., the line of latitude closest to the sun) oscillates between the two tropic lines, making one cycle per year.

Thought (and calculation) question: In Toronto, our latitude is about 44° North (hence the name of a famous local restaurant). Call it 45° to simplify the geometry. Averaged over a year (and ignoring trivial complications like possible differences in cloud cover, or the effective thickness of the atmosphere), how much *less* solar energy strikes a square metre of a flat Toronto parking lot than would strike a similar area at the equator?

Hint: You might have to get out your high school trigonometry book!

Disclaimer to hint: Actually, I have no idea whether they still teach trig in high school. Perhaps it has gone the way of punctuation rules, grammar, and cursive handwriting. Whatever. It must be on the internet somewhere.

In the thought question above, I suggested that you ignore the thickness of the atmosphere, because I wanted to focus your thinking about the simple underlying geometry of the problem of light striking a curving surface. But can we really get away with ignoring the fact that—assuming that the atmosphere is evenly distributed across the earth, which is fairly true—a photon heading for Toronto makes a longer transit through air than does a photon bound for the equator? (You can see this is true by making a diagram with concentric circles using your high-school compass [assuming they still use *those!*]. Alternatively, you could calculate *how much longer a transit it makes* by the same sort of trigonometry you brought to bear on the simple thought question.) If you do the calculation, you will see that the difference in atmospheric transit length is non-trivial. Nevertheless, it turns out that we *can* basically get away with ignoring it, at least for simple, approximate calculations. Why? Because most of the energy reaching the earth from the sun is in the spectral range of *light*, and those wavelengths pass through air without giving up much of their energy to the air molecules. Light is hardly absorbed by air. *You knew that already, of course; that's why we consider air transparent.*

How, then, does the incoming solar light energy transfer so much heat to the earth? That transfer happens when light hits surfaces *other than air*: solid surfaces, especially dark ones, or water. When photons hit those surfaces, they are absorbed and reradiated at longer, *infrared* (IR) wavelengths. Light is converted to heat. IR radiation, unlike light, is absorbed by the atmosphere. Therefore, solar energy heats the earth's surface, and then the surface heats the air *near the surface*. Therefore, solar input paradoxically heats the air at the *bottom* of the atmosphere, not the air at the top that is closest to the sun. And this heating is strongest near the equator.

Atmospheric circulation

The delivery of solar-derived heat to the bottom of the atmosphere, rather than the top, has tremendous consequences. Imagine a pot of water. If you place a heat source *above* the pot, pointing down at the water, nothing violent happens. The water at the surface would warm up, of course, and start to evaporate at a higher rate. Eventually, all of the water would be transformed to vapour, but the process would be a placid one. Things are different if you apply heat to the bottom of the pot, as with a conventional stove burner. Now you see bubbles rising, as water near the bottom turns into vapour. As you continue the heating, the water is set into increasingly violent motion as a simmer turns into a *boil*, the boil turns into a *rolling boil* (ideal for pasta!), and the rolling boil eventually *boils over*. This circulation happens because heating *from the bottom* makes the lower layers of water less dense. Packets of hotter, low-density water are more *buoyant* than the colder water above them, so they are propelled upward and the colder water sinks, in the process called *convection*. Both water and air are fluids whose buoyancies depend on temperature. Therefore, solar heating of the bottom of the atmosphere sets up *atmospheric circulation* patterns that are quite analogous to the rolling turnover that we see in a boiling pot of spaghetti water.

As you might imagine, however, atmospheric circulation differs from the comparatively chaotic circulation of a small pot of boiling water. Yes, there are pockets of violent and chaotic overturn in the earth's atmosphere. These are manifested as immensely energetic storm systems: nor'easters, thunderstorms, tornadoes, hurricanes, and typhoons. (In 2017, we have just seen how the big hurricanes Harvey and Irma were powered by exceptionally warm surface waters in the Atlantic Ocean and the Gulf of Mexico.) Such storms are powerful events, with serious consequences for both humans and the natural world. Still, such violent episodes in the earth's atmosphere are really just small, local deviations from a much smoother, predictable, and majestic pattern of circulation at the global scale. Arguably, the consequences of the large-scale patterns are even more important than those of storm systems, but we are less aware of them because they are so much more stable and predictable in character. They don't make the news, but they shape the living world.

Hadley cells

The earth's atmosphere is a relatively thin layer of gases that is pressed against the earth by gravity. The weight of a column of air compresses the air at the bottom through *atmospheric pressure*. Photon flux from the sun is concentrated at the equator (more precisely, the *solar equator*—see below), so the equatorial region of the earth's surface heats up the most. IR radiation from the heated surface in turn heats up the near-surface

atmosphere, rendering the air less dense. The reduction in density (1) causes a meteorological *low pressure zone* and (2) impels the heated air to rise above the solar equator. As the air rises, it tends to create a partial vacuum beneath it, and that suction causes surface air to be *drawn toward the solar equator* from the north and south. That new air also heats up and rises. This sets up a continuous flow—a fountain-like column of air that rises continually, like an escalator or a conveyor belt. However, the air cannot keep rising forever; it bumps up against the top of the atmosphere, and is then pushed away from the solar equator, moving to the south and the north. This flow is shown by red/blue arrows outside the globe in the climate circulation diagram (**Figure 5**, page 4).

As the air rises, it expands more because there is less atmosphere above it to compress it. Expansion of a gas causes it to cool. The relationship of temperature drop to altitude gain is theoretically described by the *adiabatic lapse rate*, which is ultimately an extension of the *ideal gas law*. The exact relationships are strongly affected by humidity, but for a 1 km rise, the air temperature drops about 5°C under saturated humidity and about 10°C in dry air. By the time it reaches the upper atmosphere, the air is no longer warm and buoyant—it is cold and heavy. That air now sinks, but because it is continually pushed to the north and south, it descends *not* at the equator, but rather at roughly 30°N and 30°S latitude. It doesn't stay cold, however, because the lapse rate relationship applies in both directions: sinking air warms up. After descending, these packets of air are pulled back toward the solar equator, where they converge and rise again. These movements establish two continuous circulation loops, each about 30° wide, completing the metaphor of an escalator: as the track under your feet carries you upward, the hidden track below goes back down. Similarly, air masses at the earth's surface tend to move toward the equator, while air masses at higher altitude are moving in the opposite direction.

Some terminology: The zone of rising, heated air is called the *intertropical convergence zone* or ITCZ. “Intertropical” is an accurate descriptive name because the solar equator moves between the two tropic lines over the course of a year. The two circulation loops are *Hadley cells*, named after a 19th-century British meteorologist who was concerned with explaining the source of the trade winds.

The Hadley cells not only set prevailing winds in motion, but also affect precipitation in profound ways. The air that comes into the ITCZ is humid, heavily laden with water vapour. As it rises and cools, much of that water vapour condenses into liquid water clouds and falls as rain,

so the equatorial tropical regions are very rainy. These are the places for tall, dripping, evergreen rainforests, teeming with life. (As you are probably aware from listening to weather reports, low-pressure weather systems are always associated with precipitation.) By the time a packet of air has reached the upper atmosphere, it has been wrung dry of most of its moisture. When that air subsequently descends at the 30° N and S latitudes, it comes down as hot, dry, and desiccating air. These latitudes chronically experience high-pressure weather systems, with the sun beating down relentlessly out of a cloudless sky. These are the places for sparsely vegetated *deserts*. As that dry air returns toward the ITCZ, it gradually picks up moisture from the ocean and the wetter tropical regions, setting up the next cycle. Perhaps ironically, the wettest and the driest places on earth both have their weather delivered by the Hadley-cell circulation.

Ferrell and polar cells

Although I have described the Hadley cells as *continuous* loops, they are not *closed* loops. When the masses of dry air descend at 30°N and 30°S, nothing predestines that air to head back toward the intertropical convergence. In fact, it is pushed indiscriminately both northward and southward. The flows that move toward the poles initiate a second pair of "conveyor belts," the *Ferrell* or *mid-latitude cells*. The Ferrell cells are not as strong or consistent as the Hadley cells, but they are driven by the same processes. As the dry air in the Ferrell cells moves across the earth's surface toward the pole, it also picks up moisture, and finally tends to rise, creating another pair of rainy and snowy low-pressure zones around 60°N and 60°S. As that rising air reaches the upper atmosphere, it is shoved toward the north and south by the continuous upward flow from beneath (just as happens at the ITCZ). The high-level flows *toward the equator* close the Ferrell loops; the flows *toward the poles* set up a third pair of circulation loops, the *polar cells*. The polar cells are the weakest and most diffuse of the three.

Prevailing winds

At different places on the earth's surface, winds tend to blow from characteristic directions and with greater or lesser force and reliability. For obvious reasons, humans became acutely aware of these patterns during the age of sailing ships. Before the age of exploration, people experienced wind as a purely local phenomenon, but maritime pioneers necessarily had to adopt global views. As with most scientific enterprises, human understanding of winds proceeded first by compiling observations and looking for consistent patterns within the observations. Then, theory and modeling came into play as tools to understand *causal mechanisms* and to allow *predictions* (Hadley and Ferrell fit into this tradition). These wind patterns are absolutely critical influences on organisms and

ecosystems, partly because they redistribute heat, but mostly because they redistribute water (as vapour) from oceans to continents. Without that redistribution of water, the continents would be barren deserts. We need to understand how this works.

By convention, winds are named by the direction they come from, so a *north wind* is one that blows *from north to south*. For simplicity, I'll restrict the following treatment to the northern hemisphere; effects in the southern hemisphere are essentially a mirror image. As described above, the six-cell circulation pattern imparts northerly and southerly components to prevailing wind directions: between 0° and 30°N , the Hadley cells push air from north to south; between 30°N and 60°N , the Ferrell cells push air in the opposite direction; and between 60°N and the North Pole, the polar cells impart a flow to the south. (We get bitter cold weather in Toronto when the polar cell bulges southward, bringing that frigid arctic air flowing from the north.)

Westerlies, easterlies, and the Coriolis effect

So far in this explanation, the only vectors of wind directionality are north or south, but you may be aware that weather patterns in Toronto are usually influenced by *prevailing westerlies*. Our weather usually blows in from the prairies. What imparts an east-west directional tendency to wind flows? There is no actual *force* at work. Although it seems counterintuitive, our westerlies arise because the air being pushed *straight northward* by the Ferrell cell is passing over the surface of a spinning sphere. That action produces a *twist* of the wind vectors with respect to the earth's surface—the *Coriolis effect*, which is sometimes called a *pseudoforce*. To understand how this works, first understand that the earth is spinning rapidly on its axis, and that the atmosphere near the earth spins with it. Specifically, a point at the equator is travelling from west to east at about 1670 km/hr. As you proceed from the equator to the North Pole, however, the speed of the earth's rotating surface slows down because the earth's diameter shrinks. Under these conditions, consider a packet of air over Omaha, at about 40°N , in the centre of the U.S. At that latitude, it will be under the influence of the Ferrell circulation, which will give the packet a northward impetus *in the direction of* Winnipeg, which lies directly north of Omaha at about 50°N . In due course, the packet will reach the latitude of 50°N , but it will *not* hit Winnipeg. Why? As the packet leaves Omaha, it is moving very fast to the east—as fast as Omaha is. Winnipeg is also moving east, but much more slowly than the packet of air we are following. Therefore, the air *will get ahead of* Winnipeg—it will probably cross 50°N somewhere in northern Ontario on its way to James Bay. If you draw a line from where the air started in Omaha to its destination in Nowheresville, Ontario, that line

does not run from south to north but rather from southwest to northeast. That is to say, the wind direction in this latitudinal range acquires a *westerly component* caused by the Coriolis effect, in addition to the *south-easterly component* caused by the Ferrell circulation. You should realize that this is like a vector-sum problem in physics.

As shown in **Figure 5** (page 4), the Coriolis effect induces the opposite twist in the latitudinal belts between the equator and 30°N or 30°S. Here the air is moving toward the equator, so air packets fall behind their apparent target rather than getting ahead. Therefore, these zones are subject to *prevailing easterlies*. Those winds were so important to maritime commerce in the sailing era that they became known as the *trade winds*. Ships from Europe bound for the Cape of Good Hope at the southern tip of Africa would typically slant westward across the Atlantic all the way to South America, taking advantage of the trade winds and getting as far south as practical, so as to take advantage of the prevailing westerlies to push them back across the South Atlantic to Africa. It was not all “smooth sailing,” however. The prevailing winds are strongest at the latitudes in the *middles* of the atmospheric cells, roughly at 15° and 45°. There, the air is primarily being pushed *horizontally* across the earth’s surface, producing consistent winds. At the latitudes where air packets are mainly going upward (0° and 60°) or coming down (30°), there is little horizontal wind, and it is very flukey. Sailors justifiably feared these areas because they could get becalmed for weeks, and these dangerous zones acquired special names. The windless equatorial area became the “*doldrums*,” and the windless areas at 30° N and S became the “*horse latitudes*.” “Being in the doldrums” has become a psychological term for listless depression. Sailors encountered the opposite problem—way too much wind!—around 45° S, where the prevailing westerlies grew so fierce that they became known as the *Roaring Forties*. These winds are so powerful and relentless because there are no significant land masses in the vast southern oceans. (Well, OK, apologies to sensitive Tasmanians, Kiwis, and Tierra del Fuegans. But, still.) The winds whip around and around the globe unimpeded. In the North forties, the continents interrupt and dissipate the flow. We don’t speak of a northern version of the Roaring Forties.

Thought question. Sailing ships could get becalmed for weeks in either the doldrums or the horse latitudes. Which was the worse place to get stuck, and why? Hint: The answer does not involve wind.

Other climate peculiarities: jet streams

For reasons far too complicated to consider in first-year biology, atmospheric circulation also sets up another class of wind currents high in the atmosphere. These *jet streams* are concentrated and narrow westerlies, and they wander around in irregular fashion. Our weather is much influenced by the northern polar jet stream, which can form at the boundary between the Ferrell and Polar cells. (There is a vivid animation in Wikipedia at *jet stream*.) As the jet streams wobble around, they blur and transgress the usual boundaries between the cells. When we get an unusual storm system or heat spell in the higher latitudes, it is frequently associated with some rogue loop in the jet stream. For example, between 30 degrees and 60 degrees north latitude, we usually experience prevailing westerlies. But there is a notoriously fearsome type of winter storm called a "Nor-easter," which drives down Arctic air from the northeast. Remember, weather is not climate. Generalizations about climate patterns in certain latitudes refer to *long-term averages*.

Oceanic circulation

Since we haven't said anything about *organisms* for a while, let me insert a reminder that the biggest influences on *what organisms live where* are *temperature and precipitation*. By considering the differential heating of the globe from equator to poles—and the ensuing consequences for atmospheric circulation, we have covered much of the underlying mechanism for global variation in climate. However, we also need to touch upon *oceanic circulation patterns*. Just as air packets of different temperatures rise and fall in the gaseous atmosphere, producing directional flows, so do massive packets of water produce circulation patterns (*currents*) in the oceans (see lecture slides). Oceanic circulation is far more complicated than atmospheric circulation, however, so we won't attempt any detailed analysis in BIO120.

There are some salient points to consider, however. First, what goes on in oceans definitely affects neighbouring land masses; *land and sea are not independent systems*. For example, Great Britain and Labrador are at similar latitudes, but the first has mild winters while the second is frigid. The difference is the Gulf Stream, a strong current that bathes the British Isles in warm water from the Caribbean. The result: palm trees grow in Cornwall. Some models of climate change suggest that global warming might disrupt the Gulf Stream, sending the United Kingdom into almost subarctic conditions. That would certainly be an ironic consequence of warming.

Maritime versus continental climates

Aside from moving heat from one place to another, oceans also influence *nearby* land masses simply by providing thermal inertia. Land masses heat up in summer and cool down in winter much faster than do masses

of water. In comparison to the centres of continents, spring arrives more slowly in coastal areas, peninsulas, and islands, but summer lingers longer into the autumn. The ocean-influenced *maritime climates* are also buffered against temperature extremes. *Continental climates*, which are found more toward the centres of big land masses, lack this water-dependent buffering of temperature. Their summers are hotter and their winters more frigid. The Great Lakes are large enough to exert some maritime-climate effects. Frost-sensitive crops are frequently cultivated in areas with maritime climates because sudden freezing is less likely. The lecture slides develop some of the consequences of the continental-maritime contrast. One of the most important is that the ITCZ follows the solar equator only *approximately*. In particular, it is distorted by the great annual temperature swings produced in the large continental mass of Asia. Look carefully at the lecture slide showing where the ITCZ lies in January and June, and work through the consequences for seasonality in the tropics. Seasonality in the temperate zone is first and foremost a matter of temperature differences. In the tropics, temperatures are comparatively uniform, and seasonality is a matter of precipitation. In a few places, the ITCZ hardly moves, creating an almost uniform rainy climate throughout the year. These places have the archetypal tropical evergreen rainforest. In other parts, the ITCZ swings over a large range of latitude, producing one or two discrete rainy seasons as it passes over, separated by dry seasons when it moves away.

Thought questions. Why do some places in the tropics experience two discrete rainy seasons per year while other locations experience one? If location A has two rainy seasons 6 months apart, and location B has two rainy seasons 4 months apart, what can you say about those locations?

Ocean temperatures and precipitation

As mentioned before, the continents get their rainfall from air that gets loaded with water vapour by passing over oceans. Much more loading takes place when the ocean waters are warm rather than cold. The warm water warms the air. Molecules of liquid water are more able to evaporate into vapour from warmer water, and warmer air can retain more water vapour. Therefore, when winds carry ocean-derived water vapour onto land masses, they will bring much more potential rain if those winds have been blowing over warm water. The clearest manifestation of this effect is indicated in the lecture slides that address the question, “Where are the world’s driest deserts?” We already know that deserts are concentrated around 30°N or 30°S, but within the large extent of desert habitats, the most extreme are located on the western coasts of continents, and each one is paired with a cold offshore current

that brings an upwelling of water from the cold depths. Therefore, the primary potential sources of precipitation for these areas are all starved of water vapour. The Canary Current gives rise to the western Sahara Desert, the Benguela to the Kalahari, the California to the Mojave, and the Western Australia to various Australian systems.

Orogenic precipitation and rain shadows

Sharp differences in rainfall also arise when terrestrial topography interacts with prevailing winds. Imagine the west coast of North America, which has several mountain ranges running more-or-less north to south. Water-laden air from the Pacific is pushed into and over these mountain ranges by the prevailing westerlies characteristic of this latitudinal belt. As this air is pushed up the windward sides of the Coastal Range (the Olympic and the Cascade Ranges), it is cooled according to the lapse rate equations. As happens with the air rising in the ITCZ, the cooling causes condensation and precipitation. The Olympic Range gets so much rain that it supports one of the lushest temperate rainforests in the world, with giant firs and cedars. Coastal British Columbia gets so much snow that it hosts Whistler (lush condos, with furs and giant Ski-Doos). These extreme biological and cultural phenomena are driven by grandiose levels of *orogenic* (mountain-generated) precipitation. But that precipitation is left behind in the mountains, while the air keeps going east.

As the now-dry air descends the eastern (leeward) slopes of the Cascade Range, etc., it warms up from compression (lapse rate applies, as always) and from friction. This is now dry and desiccating air. It's similar in quality to the descending Hadley-cell air that creates deserts at 30°, but is a more local phenomenon. The rain-starved region that results is called a *rain shadow*, with xeric vegetation that ranges from short-grass prairie to near-desert. No giant cedars on *this side* of the range!

As this dry air continues moving eastward across the dry intermountain west, it can pick up some moisture from the vegetation and from additional air flows from the north or south. But before long, it bumps up against the Rocky Mountains, and the cycle of water extraction is repeated. Because the Rockies create such a long north-south rampart, almost the whole of North America is a rain shadow in their lee (see lecture slides). Much of the hot, dry west and southwest is watered by rivers, not by local rain—and the rivers have water only because they are fed by meltwater from the winter snows that fall in the Rockies. In this part of the world, and all other dry areas, permanent rivers support narrow bands of richer, taller vegetation along their banks. This is called *riparian* vegetation, or (in the dry tropics) *gallery forest*. In North America, the

quintessential riparian trees are cottonwoods. If you see a line of them in the distance, winding across an otherwise treeless landscape, you know there's a river there. When I drive across the country on I-80 each summer, I pass a section of Nebraska where trees were so scarce that the early European settlers made their fenceposts out of stone instead of wood.

Thought question. Because winds are often more consistent on oceans than on land, oceanic islands with mountains often show particularly stark rain shadows. On Hawai'i, for example, one end is mossy dripping rainforest while the other (the Kona Coast) is dry grassland with beef grazing and coffee plantations. Is the Kona Coast on the east end or the west end of the island?

Exposure

Yet another strong effect of topography arises from differential exposure to the sun. In the northern hemisphere, south-facing hillsides face the direct, parching rays of the sun for most of the day; north-facing slopes are mostly shaded, cooler, and more moist. Consequently, the vegetation on slopes with different aspects usually differs sharply.

Part 5. Physiological ecology as a way of thinking about organisms

Living entities can be seen in different lights. In the biological sciences, we tend toward reductionist viewpoints that focus on tangible physico-chemical attributes and natural—as opposed to supernatural—causes. We stay within the realm of direct evidence. Even within that scope, however, we can look at an organism in different ways. For example, we can view it as a *bag of many juices*: a container in which stupendously complicated chemical reactions take place in aqueous solution. We can focus on the subset of those reactions that direct the development of complex adult forms from unicellular zygotes (development), or on the subset that carries out the metabolic processing of energy and materials to fuel and supply all of that growth, maintenance, and reproduction (physiology).

Alternatively, we can view an organism as a *library of information*, bearing stupendously detailed sets of genetic instructions written in DNA. That information has two special aspects. First, it is necessary and sufficient to direct the assembly of a complex organism—an astonishing property! Second, it is heritable, and thereby connects that organism to other organisms in a family tree of relatedness. Informationally viewed, each of us is a twig on that tree, as are all other organisms. That is to say, we have a *phylogenetic position* on that tree, with understandable relationships to others that can be inferred from the similarities and differences in our DNA. That information lets us classify organisms in terms of their relatedness (essentially, how similar their DNA is) into a *taxonomy* that has analogous characteristics to a pedigree.

So, are you a bag of juices, a book of code, or something more? Well, we set aside the “something more” question in BIO120. That is a qualitatively different question that needs a different conversation in a different venue. But you are most definitely both a bag of juices *and* a book of code. In fact, you couldn’t be a functioning bag of juices if you weren’t a readable book of code, and you couldn’t be a readable book of code if you weren’t a functioning bag of juices. And you share those spectacular properties with all other organisms! The evolution portion of this course has developed the “book-of-code,” genetic and phylogenetic viewpoint. Here, under the topic of *physiological ecology* or *autecology*, we emphasize the “bag-of-juices” aspects, and how those aspects relate to the environment.

How environmental factors determine distribution and abundance

All organisms have restricted spatial distributions. If environmental factors in a particular habitat are too harsh for a particular species, it will not be able to persist there. I mention this fundamental notion briefly at the beginning of my first lecture, then return to it in my second lecture. We classify environmental factors in a few different ways. One distinction is between *abiotic* factors and *biotic* factors. Abiotic factors are manifestations of the non-living, physico-chemical world, whereas biotic factors arise from the actions of other organisms. For a simple example of a biotic effect on a species distribution, consider that coyotes are almost never abundant in areas with large populations of wolves. The wolves are more powerful; they hunt the coyotes down and kill them—a stark but straightforward biotic interaction.

I will start, however, by discussing abiotic factors. These are subdivided into two categories, *conditions* and *resources*. Resources are necessary physical entities that organisms use up. They can be depleted. Examples could be water, chemical nutrients, and space. In contrast, conditions are physical states that cannot be depleted, such as temperature or pH. Factors that are most important in determining whether a species can or cannot persist in an area are called *limiting factors*. At the broadest possible level, the two factors that are most likely to limit the distributions of terrestrial species are temperature (a condition) and water (a resource). Consider these “The Big Two.” For terrestrial organisms, water availability ultimately depends on precipitation, so weather and climate are the most important determinants of what sorts of organisms are found in different parts of the world. Because of this, I spend much of my first lecture outlining how climatic variation comes about. (I also briefly list other common limitations other than temperature and water. These will be discussed later in the course.)

Of course, organisms are affected simultaneously by all of the factors in the environment, but we frequently choose to focus on only one factor at a time. The standard concept here is of an *environmental gradient*. The simplest sort of gradient would occur along a line drawn across a real habitat. For example, if you walk from the shore of a lake toward higher ground, you would be walking along a gradient of soil moisture. At the lake’s edge, the soil would be saturated with water, but it would become drier as you moved farther from the water source. Suppose you laid out a 1-m wide strip to walk along, and counted the number of plants in each meter of the strip. Your sample strip would be called a *transect* (so called because it *cuts across* ecological variation), and you would expect it to represent a gradient of soil moisture. If you identified and counted the numbers of plants in each 1-m square, you might expect some water-

loving species to be most common near the lake edge, and there might be other plant species that thrive best in the drier soils farther from the lake. We would describe those species as “sorting themselves out” along the gradient in patterns that reveal their differential sensitivity to soil moisture.

There is a different way to construct a gradient of samples, slightly more complicated but much more commonly used in ecological studies. Suppose you locate a large number of sample points scattered randomly throughout a lake valley. At each point, you set up a 1-m square sampling plot. In addition to conducting a census of the organisms in each plot, you also measure the soil moisture content. Then, in the computer, you sort the sample plots from lowest to highest soil moisture. This arranges the samples along a gradient of soil moisture, and you can look to see how different species sort themselves out along this gradient. (Variations on this practice formed the heart of plant community ecology, and I will lecture about this later in the course.) Of course, any factor can be treated as a gradient in this way: temperature, pH, salinity, organic content of soil, altitude above sea level, etc. Altitude (or elevation) has been used as a gradient in many studies of the distribution and abundance of organisms in mountainous (montane) regions, and I will lecture about these applications. Altitude is considered a complex gradient, because many simpler factors vary with altitude in nature: temperature, amount of precipitation (“The Big Two,” remember?), type of precipitation (snow or rain), partial pressure of oxygen, depth of winter snowpack, date of spring melt, length of growing season, and others. Because many of these factors are critical to the success or failure of organisms, most species in montane regions are restricted to well-defined ranges of elevation, tempered by exposure.

For now, the essential point is that any particular species is likely to be restricted to only a portion of an ecological gradient, if the factor that varies along the gradient is a limiting factor. That portion constitutes the *range of tolerance* for that particular factor, and it can also be considered as defining part of the niche of that species. Ranges of tolerance are classically graphed as curves that show how an organism’s ability to function changes along the gradient. In abstract depictions, they are usually represented by bell-shaped curves, although real tolerance curves may have very different shapes. Still, if the gradient is long enough, there is usually a peak at some point, where the environment is optimal, or best-suited for the organism to thrive. At increasing distances from their optimum, organisms find the environment increasingly stressful. First, they become unable to grow well enough to reproduce;

then they become unable to grow at all, and finally they become unable to live—think of these as the death zones. (See lecture slides.)

Thought question: In some regions of the boreal forest, the two most abundant tree species are black spruce and white spruce. If you graph the abundance of these two species across a gradient of soil moisture, white spruce has a typical, unimodal abundance curve with its peak in the middle of the range. (The special names for soil moisture levels are hydric (wet, waterlogged), xeric (dry), and mesic (intermediate between hydric and xeric).) Black spruce, however, has a bimodal abundance curve: it is more common at the xeric and hydric ends of the gradient, and less common in mesic sites. It seems hard to explain a bimodal gradient in terms of the physiology of a species. Can you suggest other explanations for the distribution of black spruce?

Why are temperature and water so important?

This comes back to the bag-of-juices metaphor for organismal function. The processes that permit life are essentially chemical reactions that occur in aqueous solutions. You should know from chemistry that reaction rates depend strongly on *temperature* and on the concentrations of the reactants. In very cold conditions, molecules move so slowly that reactions essentially come to a halt. Temperature dependence characterizes even the simplest inorganic reactions. It becomes even more important in organisms, because many of the most important reactions are catalyzed by *enzymes*. Enzymes are proteins whose ability to catalyze a reaction depends on the way that the protein molecule is folded or configured. At high temperatures, proteins *denature*. They lose their characteristic shapes and their functionalities. To some extent, natural selection can produce heat-resistant enzymes, but that resistance is limited. Because of the fundamental temperature dependence of chemical reactions, all forms of life are limited by extremes of cold and heat.

Water is important because it affects the concentrations of chemical reactants, but especially because cells and tissues depend on *membranes* to compartmentalize chemical processes and reactants. Proper functioning depends on *osmotic balance*. If cells get too dry, the concentrations of dissolved salts increase. Chemical reactions are slowed and changed. Eventually they grind to a halt; salts can even precipitate out of solution and crystallize. If too much water enters cells, reactants get diluted and fail to combine as needed.

Now, are organisms characteristically in danger of overheating, over cooling, drying out, or getting waterlogged? Absolutely! The physico-chemical reason is simple, and has three parts: first, environments typi-

cally contain a far broader range of physical conditions than the much narrower ranges of tolerance that characterize organisms. That is, *death zones are out there*. Second, things tend to equilibrate. Objects, whether organisms or inanimate objects like rocks, will tend to reach the same temperature as their environment. In the hot sun, things heat up. The same is true for water content. In a dry environment, objects will lose water and become saltier. Third, environments are much larger than organisms, so equilibration is asymmetrical. If you put a warm rabbit in a freezing Arctic environment, heat moves from the rabbit to the environment, making the rabbit colder and the environment warmer. Because the rabbit is such a tiny entity in the frozen immensity of its environment, its heat loss has a dramatic effect on its own temperature. But the tiny bit of warmth that it transfers to the environment is completely trivial, and could only be measured by incredibly sensitive instruments. *Generally, organisms are intrinsically vulnerable—environments are intrinsically implacable.*

To stay alive, organisms must be able to combat this tendency to match the environment. They must be able to keep their internal states constant (or at least within narrow bounds), a process known as *homeostasis* (“staying the same”). The two most important categories of homeostatic mechanisms are involved with maintaining temperature (*thermoregulation*) and saltiness (*osmoregulation*). A rabbit on a freezing Arctic plain must replace the heat it loses to the surrounding Big Chill by exothermically oxidizing carbon compounds from the food it eats.

Organisms vary hugely in their capabilities for thermoregulation and osmoregulation. Much of this depends on body size and shape. As I will develop in lecture, the most important aspect of body structure is the ratio of surface area to volume (SA/V). Consider a tiny, aquatic, single-celled *Paramecium* in a freshwater pond. Because it is so small, its SA/V ratio is high: any minuscule amount of heat that it could generate in its tiny body would almost instantly be sucked away by the overwhelming mass of water. Pond temperatures don’t reach extremes, and these little protozoa can get away with being *thermoconformers* rather than *thermoregulators*. But their need for osmoregulation is imperative, because water is continually seeking to lethally dilute their protoplasmic bag of juices. If the flow can’t be countered, the cell bursts. If you watch a *Paramecium* under the microscope (or on YouTube, of course), you will see its contractile vacuole continuously filling up with water (like a water balloon) and then squeezing to expel the water back into the pond. That is its mechanism for osmoregulation—a lovely, visible illustration of how the struggle for existence is waged by a comparatively simple life

form. Of course, this squeezing-out process requires expending energy to counteract the unrelenting passive inward pressure of the water. That expenditure of energy makes it what we call *active* regulation.

The organisms most capable of homeostasis have large bodies and complicated metabolisms. In lecture, I will concentrate on thermoregulation by the vertebrate animals that have brought that art to its highest form: birds and mammals, the so-called “warm-blooded animals.” These animals exhibit very active regulation of heat balance to keep their body temperatures almost constant, despite great fluctuations in ambient temperature. In contrast, the “cold-blooded” animals (fish, amphibians, reptiles and virtually all invertebrates) tend to be *thermoconformers* whose body temperatures more closely track ambient temperatures. There are more precise terms for this dichotomy: “*homeothermy versus poikilothermy*” stresses the *constancy* of body temperature, as opposed to variability. “*Endothermy versus ectothermy*” emphasizes that temperature is primarily determined by physiological processes acting *within the body*, as opposed to being determined by the external environment. **Figure 6** (page 5) shows an exception: even insects can be thermoregulators if they are large enough, have access to high-energy food, and have insulation. Bumble bees incubate their offspring much as birds do, and they can forage in air temperatures as low as 5°C if necessary.

Thought (and calculation) questions: Different species of Paramecium range from 50 to 330 microns in length. Assuming for simplicity that they are spherical, what would be the SA/V for the smallest and largest species? Now, in fact, they are **not** spherical, but have elongated forms ranging from football to cigar shapes. How would your SA/V estimates change if you took their actual shapes into account? How would you expect the activity of the contractile vacuole to differ between large and small species? Imagine a Paramecium that is drifting down a river toward the ocean, with the water getting increasingly saline. How would you expect the action of the contractile vacuole to be changing? (If thinking about this kind of question interests you, you might be interested in a nice introduction to ecological modeling by my colleague from Berkeley, John Harte: Consider a Spherical Cow. The spherical cow reference is from an old joke about theoretical physicists trying to approach real-world problems.)

Heat balance in endotherms

You can make a partial analogy to heat balance by considering your credit card statement: you put money in, you take money out, and if more goes out than comes in, trouble ensues. For animals, the balancing act is trickier, because they also suffer if too much heat comes *in*—that’s not a problem for a credit account. Also, if your payments don’t keep

up with your withdrawals, you just rack up debt and pay more interest. If an animal develops too much of a heat imbalance, it pays with its life.

Component processes in heat transfer

Several processes contribute to heat transfer between an animal and its environment, as I will outline in lecture. *Conduction* is the direct transfer of heat between two bodies that are in contact. If there is a temperature differential, heat will flow from the warmer object to the cooler. Given enough time, the two objects will *equilibrate* at the same temperature. If you stand on frozen ground, heat moves from your feet into the ground. You lose heat, your feet feel cold, the ground beneath your feet heats up a small amount. Simple. *Convection* is heat transfer that is facilitated by a moving fluid—typically air or water. If you stand in an icy flowing stream, the cold water not only extracts heat from your feet in a conduction-like process, but the current sweeps the slightly heated water away, so local equilibration is minimal. Convective flows therefore tend to increase rates of heat transfer. Water is a *much* more effective medium for heat exchange than air is. Although they are both fluids that allow convection, water has a higher *specific heat* than any other material commonly found in ecosystems. That means that more energy must be transferred to change its temperature; in consequence, cold water sucks heat out of a warm animal *very* effectively. An air temperature of 15°C is just chilly, but being immersed in 15°C water will kill a human within a few hours.

Water is also special in that it requires much heat input to change its state from liquid to gas, *i.e.*, it has a high *heat of evaporation*. This means that *evaporation from a moist surface* is a very effective means of cooling that surface. We know from personal experience on hot days that we feel cooler with a fan blowing on our skin; if we moisten the skin, the cooling is substantially enhanced. The combination of convective and evaporative cooling is potent.

All of the above components of heat transfer involve molecules bashing into each other and transmitting their kinetic energy. The final type, *radiative heat transfer*, does not. All bodies emit electromagnetic *thermal radiation* depending on their temperatures. Like light leaving a light bulb, this energy zips out in straight lines, in all directions. Indeed, light waves do transfer some heat, but most radiative heat transfer involves longer, invisible wavelengths in the infrared. The net change in heat that you experience through radiation depends on the net difference between the energy being radiated outward from you and the amount being radiated toward you by all the objects in the environment. This is why being outside on a clear night can make you feel so cold; almost no

radiant heat comes down from the black void of space. If you have sat around a campfire on a cold night, you have gratefully felt the warmth of the fire. Was the fire heating you by raising the air temperature in your vicinity? If so, your front and back sides would have benefited equally. But when you sit facing a fire, your face is warm and your back is cold. Therefore, it is radiant heat that is conveying your one-sided comfort.

Homeostasis and heat balance

Having discussed the various modes of heat *transfer*, we need to focus more on the resulting heat *balance*: how do organisms actually manage the gains and losses of heat to keep their body temperature within suitable ranges? Remember that the active vertebrates we are considering generate considerable amounts of heat from their internal metabolic functions, especially muscle contractions. Shivering is an effective way to increase internal heat production. In a cold environment, animals need to conserve that heat or generate more of it, but in a hot environment, they need to dump it or make less. The most basic adaptations—if not necessarily the most effective—for controlling these balances concern the size and shape of the animals. Let's examine two venerable rules of thumb.

Size: Bergmann's rule (promulgated by Carl Bergmann, 1847)

If one examines a number of closely related homeothermic species that occur in habitats with different temperatures, there is a tendency for the species from cold environments to have larger body sizes. Because larger bodies translate to lower SA/V ratios, they retain heat better and lose less of it to their cold surroundings. In a tropical environment, an animal the size of a polar bear would have a hard time shedding heat, and tropical bears are gratifyingly small. There are enough exceptions to Bergmann's rule that some have dismissed it, but recent reviews suggest that it has enough validity to be retained (and may even apply to some insects). Its relationship to SA/V is satisfyingly simple to comprehend. In lecture, I will discuss some exceptions, both apparent and real.

Shape: Allen's rule

Thirty years after Bergmann, J.A. Allen described a second pattern, also explained by SA/V considerations. He claimed that animals from hotter environments had longer and thinner appendages: legs, tails, ears, sometimes horns or frills. That difference in shape will increase the surface area, given a constant volume. As lecture slides will demonstrate, rabbits are a group in which this trend is nicely shown. The comically large, almost paper-thin ears of rabbits are classic cases of maximizing surface area. (Also review my comments on small-eared pikas in Part 2, above.) Aside from affecting SA/V relationships in purely geometrical ways, extremities can be further specialized for dumping heat from the

body core by managing blood circulation. Rabbit ears are heavily vascularized, and the flow of blood to them is *regulated*. Sphincter muscles around arterioles are especially effective in regulating how much blood is pumped into ears, for example. An overheated rabbit that needs to reduce its body temperature can pump hot blood to the ears, where the large surface area allows quick equilibration. Rabbits can also position their ears to increase or decrease heat exchange.

Interpreting an animal's extremities according to Allen's rule can be tricky, though, because Bergmann's rule is likely to be in operation at the same time. A rhinoceros-sized animal cannot have the proportions of a gazelle. Its legs would snap like toothpicks. Why? Because the strength of a leg varies with its cross-sectional area, which increases as the *square* of linear dimensions. The mass that the leg has to support will vary with the *cube* of linear dimension. So, if a polar mammal needs a larger body in accordance with Bergmann's rule, it will need stronger, stockier legs just to support its greater weight. Are we justified in citing those stocky legs as *adaptations for heat conservation*, in accordance with Allen's rule? Maybe Bergmann's Rule is the driving factor, and the stocky legs might reflect nothing more than the mechanical *constraint* that heavier animals need proportionally stronger legs. Problems like this pop up all the time when one is trying to use adaptive reasoning. How do you work around the problem that legs have multiple functions? One way is to sidestep ambiguous characteristics and concentrate instead on those that should be less constrained. In this case, you could look at ears instead of legs. There is no mechanical reason why an animal would need smaller ears just because selection has given it a larger body. Therefore, if an arctic species has smaller ears, it's more likely to be a direct result of selection for heat conservation, rather than a side effect of some other response. Mammalian ears should be more "free" to register the evolutionary effects of Allen's rule than legs might be. Birds' wings, in contrast, might be expected to be *highly* constrained by their need to function in flight. One would not expect to find much relationship between wing length and environmental temperature.

Insulation

One reason for questioning the primacy of Bergmann's rule is that the same heat-conservation advantages that can be achieved by growing larger could also be achieved by adding insulation. Adding insulation in cold climates would seem to be energetically cheaper and evolutionarily "easier" than growing a larger body, with all of the redesign that would be required. Lecture slides will illustrate some different modes of insulation and some extreme cases.

Homeothermic vertebrates achieve insulation with subcutaneous layers of body fat, or as layers of fur or feathers. In the case of fur and feathers, the real insulation is provided by dead air spaces within the material. Air is a great insulator if it is prevented from moving. The real role of the fur/feathers is to trap the air and thereby prevent convective flow. Within the layer of insulation, a *temperature gradient* is set up and maintained; temperatures are cold near the outside of the fur, but warm next to the skin. Different insulation materials are most appropriate in different environments. Feathers and fur share some useful characteristics. Both comprise thousands of modular structures that can be replaced as they wear out, so the insulating layer doesn't degrade. Also, animals can grow insulation that is appropriate to colder or warmer times of year. If you have a pet cat or (especially) a dog, you are doubtless familiar with seasonal shedding. Insulation thickness can also be adjusted on much shorter time scales. Mammals can increase fur thickness by contracting the *arrector pili* muscles in each follicle, causing hairs to stand up. A bird can similarly lift its outer *contour feathers* away from its body, creating a dead-air space underneath that is filled by very fine but springy *down feathers*. Down is extremely light, but sufficient to block convective airflow. The ability of birds to instantly puff themselves up or sleek themselves down is critical, because they need to be sleek and streamlined to reduce air resistance while flying.

The combination of contour feathers and down is so effective that most birds can withstand great cold as long as they have access to good quality food. For example, Rufous Hummingbirds are the second smallest bird species in North America. They breed in the Pacific Northwest and usually migrate to Mexico, where flowers are available all winter. Because of their huge SA/V, they would experience tremendous heat loss in a cold environment. However, every year some of them get confused on migration and end up in cold places. A friend of mine kept a hummingbird feeder going all winter for a single lost Rufous that ended up in snowy North Carolina. The point is that the insulation system is spectacularly effective *if* there is enough high-energy food to run a bird's fast metabolism. If you own a thick down jacket, you know that it is unmatched in warmth per unit weight.

As previously mentioned, cold water is brutally effective at extracting heat from a warm animal. Land-based, part-time swimmers like penguins and otters can make use of regular dead-air insulation; they are able to keep their fur or feathers temporarily waterproof by oil secretions and preening. For larger mammals that are continually immersed (e.g., dolphins, whales), fur would become hopelessly waterlogged. It

would also add drag that would impede swimming. These animals depend on a thick layer of subcutaneous blubber (see lecture slides), which serves as not only as insulation but also as long-term storage of food energy. Their skin temperatures would be scarcely greater than the cold water around them, but their principal organ systems and musculature stay warm inside the encircling jacket of fat.

Countercurrent circulation in extremities

We have discussed how an overheated rabbit can usefully dump core heat to the environment by increasing the blood circulation to its ears. In a cold environment, of course, that loss of heat would be a liability. We would expect that circulation would be reduced in such circumstances, but we find a further anatomical subtlety in animals whose appendages are exposed to cold. Imagine a whale's flipper. To fulfill its primary function—powerful swimming strokes—it must (1) stick out into the water, (2) have a large surface area, and (3) be thin. These *mechanical constraints* all make the flipper a terrible liability with respect to heat loss. However, this liability can be reduced by an arrangement of blood vessels called *countercurrent circulation*. This simply entails *direct contact* between the arteries that send warm blood out to the flipper and the veins that bring cooled blood back to the body. With these vessels closely appressed, heat exchange can occur between the arterial and the venous blood. This means that the cooled returning venous blood captures warmth from the outgoing arterial blood before that warmth can be lost to the environment. Instead, the returning venous blood carries that heat right back into the animal's core. Furthermore, because the blood flows are going in opposite directions (countercurrent), there is a *continuous temperature gradient* between the two vessels (see schematic diagram in lecture slides). Therefore, the transfer of heat continues along the entire appressed length of the paired vessels.

Evaporative cooling

Because of water's great *heat of evaporation*, exposing moist surfaces to air flow is a very effective way to cool a body. Lecture slides will give examples of using both internal and external water for this purpose. Sweating works best with bare skin, so it is particularly effective for humans. Mammals that are mostly covered with fur, such as familiar cats and dogs, sweat from exposed skin such as paw pads. They also achieve evaporative cooling through *panting*, which draws air over wet surfaces of the tongue, pharynx, and lungs. Birds achieve most of their evaporative cooling through their respiratory system, which includes a system of air sacs that provides a more continuous flow, rather than the less efficient in-and-out airflow of mammals.

Just as many homeotherms can survive extreme cold if they have sufficient food, so can most of them withstand extreme heat if they have access to enough water. This means that heat becomes a pressing challenge only in desert environments where free water is hard to come by. For humans at least, deserts are deceptive. Because the air is dry, sweat evaporates so quickly that you can be unaware of how much water you are losing. The rule of thumb is that you need to carry a gallon per day if you are hiking in desert. That's a heavy load! In the deserts of the American Southwest, naïve tourists and desperate economic migrants die each year because they simply did not carry enough water.

Animals adapted to deserts have spectacular adaptations for conserving water or—in extreme cases—for doing without it. I'll mention in lecture the case of kangaroo rats, which are seed-eating rodents of the American Southwest that can survive without ever drinking liquid water. They hop around on their hind feet like miniature versions of their Australian namesakes. Their erect posture and bipedal habit reduce the amount of conductive heat gained from the hot desert sands. The seeds they eat contain very small amounts of water, but their real trick is to subsist on "*metabolic water*" produced by oxidizing the dry food they eat. Kangaroo rats also have extremely efficient kidneys that produce highly concentrated urine, so they can eliminate nitrogenous waste with a minimum of water loss. These physiological processes are complemented by behaviour. Like many desert animals, k-rats are nocturnal, so they are not exposed to the harshest heat and desiccating conditions. They hop around at night, packing large loads of seeds into cheek pouches for carrying home. They spend the day in underground burrows that are relatively cool and moist. Because they store large *caches of seeds* in the burrows where they sleep, they benefit from a particularly subtle water-conservation mechanism: seeds are *hygroscopic*, so as the k-rats exhale air moistened in their lungs, the seeds catch and retain some of that water. When the rodents eat the seeds, they get their water back. So, the imperative of water conservation is stamped deeply into these animals' *anatomies, physiologies, and behaviours*. Environments that would be dead zones for humans and many other animals are comfortable homes for species with appropriate adaptations. Of course, "comfort" is a relative term in the natural world! Kangaroo rats may have evolutionarily solved their immediate water problems, but they must compete with other animals—birds, other rodents, and harvester ants—for those essential seeds. Their nocturnal foraging is overseen by hungry owls, and their comfy burrows are subject to random home invasions by questing rattlesnakes. The struggle for existence has many dimensions.

Multi-part thought question: Adult harp seals have short, sleek fur and very thick layers of blubber; they spend most of their time in cold water. Babies, however, are born on ice sheets, with fluffy white coats of long fur. They don't enter the water until they have grown for some time and shed their long white fur. Although they are tended and nursed by their mothers for their first couple of weeks while on the ice, they are very vulnerable to predators (and human hunters, in recent centuries) while they are on land. This extreme vulnerability means that the on-ice period of this life history requires an explanation. Try to craft a sound, ultimate explanation for why this species displays this brief terrestrial period, even though it seems maladaptive in terms of exposure to predators. Explicitly consider constraints and tradeoffs. Also consider comparative evidence. For example, if you are tempted to claim that it is impossible for marine mammals to give birth and nurse babies under water, consider the evidence that whales and dolphins do just that.

Plant physiological ecology as a tangle of constraints and tradeoffs

Typical “higher” plants—the gymnosperms and the angiosperms—differ radically from the sorts of animals we have just been discussing, but a number of the same principles and tradeoffs apply. SA/V considerations are just as important, if not more so; heat balance and water balance are critical; and *evaporative cooling is the process that ties heat regulation and water regulation together*, generating tradeoffs as in animals. The big difference is that plants are *autotrophs* that harvest energy from sunlight through photosynthesis. That essential task imposes a whole new set of constraints.

With a few exceptions like cacti, the terrestrial plants that we will be considering do their photosynthesizing in *leaves*. Recall that photosynthesis requires light energy to drive the combination of gaseous carbon dioxide and water to yield gaseous oxygen and sugar. To harvest the essential photons, leaves need to offer up *large surface areas* to sunlight. They also need surface area across which gas exchange can take place: CO₂ in, O₂ out. But exposing a large surface area to sunlight immediately puts tissues at risk of heating up to the point where enzymes denature and lose function. A thin leaf lacks the volume to give it thermal inertia, so it heats up fast. Without effective ways to dump heat, such a leaf is toast. A typical leaf’s elegant solution to overheating is to use the power of evaporative cooling. Gas exchange takes place through openings called *stomata* (“little mouths”), which can be closed if needed. In addition to the CO₂-O₂ traffic already mentioned, water from the leaf’s interior can also be evaporated through the stomata. If a plant has an uninter-

rupted supply of water, running from moist soil into the roots and up through the plant's vasculature to the leaves, and if there is enough air flow across the leaf to foster evaporation, all is well. The leaf can stay cool despite receiving full sun. If the water supply is limited, however, increasingly intricate adaptations are necessary to permit a plant to harvest sunlight without overheating.

Some of these are biochemical specializations of the photosynthetic pathways; the pathways are named for the way that CO₂ capture, or *carbon fixation*, takes place. The majority of familiar plants from our temperate zone exhibit the C3 pathway, so named because an atmospheric CO₂ molecule is first incorporated into a *three*-carbon molecule. This pathway is not especially efficient in water use, and the enzyme that captures the carbon is not specific enough: it also captures oxygen, a problem that causes wasteful *photorespiration* at high temperatures. At high temperatures and in very bright light, C3 plants are at a disadvantage; those conditions favour another mode of carbon fixation in which different microanatomy and enzymes draw CO₂ molecules into a *four*-carbon intermediate. C4 metabolism has evolved independently in numerous lineages of plants, but those most important to humans are tropical grasses, including maize and sugar cane. (Wheat and rice are C3.)

The third major pathway is abbreviated CAM, for *crassulacean acid metabolism*. It provides extreme water conservation, mostly in *succulent* plants like cacti or in epiphytes. This pathway is instructive because it constitutes a breakthrough that eliminates a seemingly unavoidable tradeoff. For C3 or C4 photosynthesis to proceed, the stomata have to be open to gas exchange while sunshine is hitting the leaf. In those circumstances, water loss through the stomata is great. In CAM plants, the plants keep most of their stomata *closed during sunny periods*, thereby minimizing water loss. By having thick succulent leaves (or photosynthesizing stems, in the case of leafless cacti) that are loaded with water, they have enough thermal inertia to resist the overheating that would occur with paper-thin leaves. How, then, do CAM plants take in the CO₂ that they need? Just as kangaroo rats go foraging for seeds at night to avoid the stress of the desert sun, so do CAM plants "forage" nocturnally for CO₂ by *opening their stomata at night*. They cannot complete the entire photosynthetic pathway without the energy of sunlight, but they can capture CO₂ and store it in vacuoles as an organic acid intermediate until the following day. When photons are available, the acid can enter the rest of the pathway for converting light energy into sugar. The plant's tissues show daily cycles of pH change, accounting for the "acid"

part of the pathway name. (In case you are curious, the “crassulacean” part comes from the pathway being first discovered in the plant family Crassulaceae, although it has also evolved independently in other families with succulent tissues, such as the Cactaceae.)

Note: If you are interested in further detail regarding these biochemical details, you can of course hunt on the web. As of August 2014, the Wikipedia entry on C4 was quite good, the one on C3 was weak, and the one on CAM has a lot of detail but has been poorly edited. These entries may well have been improved, but as always with the Internet, *caveat emptor*. If you want the real story, take EEB328 (Physiological Ecology), taught by Prof. Rowan Sage, a world authority.

Micrometeorology at the leaf surface: laminar versus turbulent flow

Because gas exchange is so important at leaf surfaces, the small-scale physics of airflow becomes important, too. A leaf that is swept by wind receives a continually refreshed supply of CO₂, and the waste product O₂ is dissipated. If the air is not moving, though, stagnant air builds up next to the leaf. This dead air gets depleted of desirable CO₂ and enriched in undesirable oxygen. It is also likely to heat up because convective evaporative cooling is reduced. All three things are detrimental. Significantly and surprisingly, a layer of stagnant air can build up on a leaf surface even if a current of air is passing across it. This happens when the surface is smooth, without bumps or ridges that the air has to detour over. When airflow is unimpeded, a stratified pattern builds up called *laminar flow*. Laminae are simply layers of air that move at different speeds. Imagine that a breeze of 5 km/hr is moving across a plant. One cm above the leaf surface, the air will be going at that speed. No surprise there. But much closer to a smooth leaf’s surface (roughly 1 mm or less), the air forms a *boundary layer* that is virtually stagnant. Friction with the leaf’s surface slows air movement to a crawl. Gas exchange within this still, stagnant boundary layer occurs only by regular molecular diffusion; it gains nothing from the wind passing above.

If the leaf surface has enough irregularities, however, smooth laminar flow turns into *turbulent flow*, and the boundary layer is broken up and freshened by eddy currents and vortices. In a jet airplane, turbulent airflow would be disastrous because the eddies cause drag and decrease lift. On a leaf, turbulence will usually be desirable. Thin, flat leaves frequently have irregular or serrated margins serve to induce turbulence. One classic example is oak trees of the deciduous forest biome. The leaves of these oaks typically have their margins broken up by indentations called sinuses. As you move from the lower leaves on a tree, which are usually shaded, to the upper leaves, which are usually ex-

posed to more light and heat, the sinuses deepen and the leaves get smaller. Both changes are consistent with the hypothesis that sinuses are involved with thermal regulation of the leaves via induction of turbulent flow. Many leaves have prominent veins that stick up above the otherwise flat surface. In addition to their other functions of stiffening the leaf and circulating water, nutrients, and photosynthate, such veins help induce turbulent air flow.

Other adaptations for desert plants

Aside from physiology and biochemistry, desert plants show morphological specializations that will be discussed in lecture. Some of the most interesting concern water harvesting and storage. Such tactics are best displayed in subtropical, seasonal deserts that get rain at one or two times of year but are bone-dry in between brief rainy seasons (remember that ITCZ passing over?). The iconic giant saguaro cactus of the Sonoran desert exemplifies the *internal water storage* strategy. For comparison, consider the somewhat dry grassland of the Midwestern prairies, where most plants produce very *deep* root systems that seek out water that is held year-round, deep in the soil. The soil acts as an *external store* or bank of moisture, from which plants make withdrawals as they need it. In the much drier Sonora, there is essentially no water in the soil except right after a rain. The saguaro produces a very extensive *shallow* root system. The function of those shallow roots is to suck up that rainwater as soon as it falls, because if they don't, some other plant will get it first. Here, the cacti themselves function as the bank. They tank up with water whenever it is available, and can then call on that *internal store* as needed. The saguaro's accordion-pleated stems allow it to expand without bursting while it is imbibing after a rain. By the end of the dry season, the cactus's girth will have noticeably shrunk.

Not all desert plants are cacti, by any means. We have mentioned desert annual plants (and will return to them when we consider life histories). These plants have a compressed life cycle in which the seeds germinate right after heavy rains start. They grow, flower, set seed, and die, all in two months or so. These plants *temporally evade* desert conditions by growing during brief interludes of non-desertlike conditions. As you might expect, they don't show many extreme anatomical or physiological adaptations to dry conditions.

Of perennial plants other than cacti, woody shrubs or treelets are common. Many of these exhibit *microphylls*, or tiny leaves. These don't expose broad areas of photosynthetic tissue to the sun, and they are frequently arranged in ways to produce turbulent airflow. For their size, they are fairly thick, which gives them a little more thermal inertia. As I

will develop in lecture, some of these shrubs are evergreen, but others take the rather drastic step of dropping their leaves entirely in the driest part of the year. Some, like the palo verde tree mentioned in lecture, produce green photosynthetic bark, so they can photosynthesize with their stems even when they have no leaves. Stems don't offer enough surface area to contribute *much* photosynthesis. In the struggle to exist in a harsh climate, though, a little bit of growth is better than nothing. Cacti (and other convergent succulents such as euphorbias) are simply plants that have taken the green stem strategy to an extreme. Their leaves have been modified into thorns, abandoning their ancestral function of photosynthesis and taking on the new function of protecting the plant from water-seeking animals. As you can imagine, a plant that adopts a policy of banking its own water supply needs to protect itself against bank robbers. (Supposedly, when the bank robber Willie Sutton was asked why he robbed banks, his immortal reply was, "Because that's where the money is." For desert organisms, water is money.)

The sclerophyll puzzle

A final example from plants will illustrate an important point in crafting adaptive explanations of ecological patterns: one characteristic can serve different functions in different habitats. *Sclerophyll* is the characteristic of having leaves that are small, thick, tough and leathery, and almost always evergreen. They often contain "essential oils," pungent-smelling and flammable. Pine needles are classic examples. Frequently, angiosperm sclerophylls have leaves that are wider than conifer needles, but they are still rather narrow and simple in shape. Check out fresh rosemary in a grocery store for an example. Although an angiosperm, rosemary plants look enough like conifers that they are pruned into conical shapes and sold around Toronto as miniature potted Christmas trees for the kitchen. Sclerophyllous leaves are in many respects the opposite of the large, thin, delicate leaves—such as maple leaves—that characterize the deciduous forest biome.

This convergent evolution of a tightly co-varying suite of leaf characters would lead you to expect that sclerophylls would come from some particular type of habitat that drove selection for this signature set of traits. Our previous discussion of the functions of leaves would suggest that this particular habitat would offer some particular combination of heat stress and water stress. In fact, the truth is more complicated. Sclerophyll prevails in at least *four* habitats that present very distinct patterns of stress: First, as semi-submerged plants in acid bogs and ponds in various temperate habitats. Second, in the cold boreal forest biome, a snowy environment with reasonable amounts of summer rain and soil moisture. Third, in very dry, "well-drained" sandy soils in warm habi-

tats ranging from the southern U.S. to Central and South America, and equivalent habitats in the Old World. Fourth, in the Mediterranean/chaparral biome, with wet but not snowy winters and very dry, hot summers. So the “Big Two” conditions range all over the map, from hot to cold and wet to dry. Can we sort out why these tough little leaves are appropriate adaptations for these very different sets of conditions?

In the boreal forest, one factor is that winters are long and the growing season is correspondingly short. This gives an advantage to evergreen trees that don’t shed their leaves for the winter. Such trees can start photosynthesizing as soon as an early spring day offers temperatures high enough, whereas a deciduous tree would have to wait until the weather got dependably warm. Then it would still have to mobilize a new set of leaves before it could conduct any photosynthesis. And a tree with larger, broad leaves would still run the risk of losing those leaves to freakishly late snowfalls in the spring (or early ones in the fall). Needle-like sclerophyll leaves are better able to shed snow, especially in combination with the tall, conical growth form of coniferous trees such as firs and spruces. The flexible branches of these trees bend downward under the weight of snow, and the smooth needles—now facing downward—shed the snow without damage. If a broadleafed tree such as an oak or a maple gets caught by an unexpected snow when it is in full leaf, the results are broken limbs.

In chaparral habitats, snow is not a problem. The hot, dry summers, however, are virtually desert-like. Therefore, the same forces that favour microphylls in deserts will be at work in chaparral. Small leaves are less likely to overheat, and therefore don’t require as much water for evaporative cooling.

When you find pockets of sclerophyllous vegetation within regions where climatic conditions generally favour broadleafed plants, the main driver is usually nutrient-poor sandy soils. For example, on Long Island and in central New Jersey, areas of “pine barrens” occur on pockets of *glacial outwash plains* where alluvial deposition has left soils of almost pure sand. The dominant trees in these pockets are sclerophyllous pines, with an understory of shrubby angiosperms that come from other families but are also sclerophylls. The extreme soils exert two stresses. First, sandy soils have little water-retaining capacity, so plants can occasionally experience desert-like droughty conditions during summer heat waves between rains. More important is the lack of soil nutrients, especially nitrogen. This favours the evergreen habit because keeping a leaf for several years is much cheaper than dropping it and replacing it every

year. In turn, the evergreen habit favours sclerophyll because leaves that will be retained for a long time have to be tough. Flimsy, broad leaves would be torn to shreds by winter ice storms, or would cause branch breakage through snow buildup.

In cool northern bog lakes, neither desert-like heat nor lack of water can pose problems, but nutrient limitation can and does. Bogs develop acidic water, with pH values as low as 4 or so. In those conditions, dead vegetation tends to be preserved from decay through a process akin to pickling. It is hard for plants to take up nitrogen. As in dry sandy soils, sclerophyll is favoured *indirectly* because being evergreen is favoured *directly* through its advantages in nutrient conservation. Bogs are also noteworthy for the high frequency of carnivorous plants (e.g., Venus's flytraps, pitcher plants, sundews, and bladderworts). By trapping insects or crustaceans, these plants acquire alternative sources of nitrogen that other bog plants cannot tap.

Acid bogs preserve more than plant material. Much of what we know of past climates comes from analyzing pollen grains preserved in layers at the bottom of bog lakes. Much of what we know of northern Europeans from the Iron Age comes from the well-preserved bodies, clothes, and stomach contents of humans who died in bogs. Rather than rotting away, their bodies became pickled and their skins tanned like leather. These bodies occasionally come to light during peat-cutting operations; most of them seem to have been victims of execution or religious sacrifice. There is a great book on this, *The Bog People*, by a Danish archaeologist with the evocative name of Glob. I'll show some of the grisly pictures in a pre-lecture loop around Halloween.

I hope that you never have to dispose of a murder victim's body, but if you do, *don't* bury it in a peat bog. You might as well have it embalmed. (Who says that BIO120 doesn't prepare students for practical, real-world situations?)

Two thought questions: I argued above that a whale's flipper has to be large and thin to achieve its primary function, and that the large/thin requirement goes on to generate secondary problems for the whale. Perhaps there is an analogy to mesophylllic plant leaves, which are also large and thin. Explore this analogy. Is it a good, deep analogy, or just a superficial one? What are the similarities and differences?

I lectured about why weasels are particularly long and thin, with short legs for their body weight. The same exaggerations of body shape are found in the dachshund breed of domestic dog. To what extent, if any, are the explanations similar for weasels and wiener dogs?

Introduction to population ecology: a taste of theory

Part 6. Population ecology

Now we move from considering individual organisms to a higher level of organization, the population. This entails a radically different style of science.

Questions in *population ecology* revolve around the number of individual organisms in a population, also called the *population size*, and usually denoted by the symbol N . Most textbooks and papers in this field conceal a dirty little secret: it is very, very hard to estimate population sizes in nature with much precision, especially for mobile animals! Part of the difficulty lies in finding and enumerating the organisms. Little animals such as insects are frequently cryptic and furtive. Even big ones are elusive. I know from radio-tracking data that I have personally been within 20 metres of a mountain lion at the Hastings Reservation in California, probably several times. They surely saw me, but I never saw them. Another hard part is deciding where the *boundaries* of the population are: which members of the species belong to a particular population and which do not? Individual mountain lions roam over huge home ranges, up to 100km^2 for males. It's obviously hard to pin down their membership in a particular interbreeding group.

Some kinds of organisms present a further difficulty: it is not always easy to distinguish individuals. This is especially true for many plants in which a single genetic individual (a genet) can subdivide itself and grow into a large clone of genetically identical plants (ramets). Sometimes these genetically identical ramets become disconnected and continue their lives as independent physiological units; sometimes they remain connected and interdependent. (Some sessile invertebrates behave similarly.) Unsurprisingly, population ecology was developed mostly with typical, mobile animals in mind.

Because of these difficulties, the field of population ecology has developed a number of distinctive features. One is a great reliance on *mathematical models and theory*. Models use the power and reliability of mathematics to compute the consequences of reasonable assumptions about population growth, even if we lack precise data. When researchers do actually study population sizes by direct enumeration, those studies tend to be carried out either on small islands or in artificial conditions in laboratory containers. In such situations, there is no uncertainty about the limits of the population, and small units are easier to search. The exceptions—those cases where scientists *do* try to estimate population sizes in continuous habitats in nature—tend to fall into two categories. The first involves pest species in agriculture or forestry or epidemiology, where the *economic costs* of population outbreaks provide a compelling

rationale for research. In such cases, one doesn't need to know exact numbers, just whether there is going to be a big year or a small one. Such predictions can allow appropriate control measures to be put in place before it's too late. For example, do we need to attack Zika vectoring mosquitos or not? The second involves dwindling populations of *endangered species*, typically vertebrates. Here, the driving motive is likely to arise from *legal responsibilities* to devise management plans to stave off extinction. Here, exact numbers do matter. Usually, the biological questions are stubbornly intertwined with controversial political issues, such as preserving habitat to preserve species versus developing it to make money.

Direct empirical studies tend to be very idiosyncratic. Their outcomes depend on particular details of the biology of the species and the landscapes in which it lives. Although every such study can provide useful insights, they provide the sorts of lessons that can be drawn from case histories, rather than generalizations of broad applicability. In population ecology, the only generality resides in the theory. Because BIO120 is focused on broad generalizations, I will emphasize the simple mathematical models that have given the field its distinctive intellectual character. This orientation also provides more value for the majority of BIO120 students, who are unlikely to become practicing population ecologists, but who are certain to encounter mathematical models in other branches of the sciences. Learning the characteristics of population models, along with their strengths and weaknesses, will help you confront other models—in physiology, epidemiology, or genetics—with more understanding and sophistication. Think, for example, how often you have heard something described as “growing exponentially.” With some understanding of the relevant equations, you will be able to determine whether that is an accurate description of the process in question, or just some thoughtlessly misused catchphrase.

“As time goes by”— modeling population changes through time

Population modeling in ecology (and modeling in general) can be broken down into two mathematical styles. One is the more classical approach, which was well established by the 1930’s. It comprises analytical, *deterministic* models that use the tools of calculus—differential equations, in particular—to make predictions about population changes through time. This approach was adopted because calculus was really the only practical tool available at that time to consider rates of change. If you could write an equation of the form “ $dN/dt = \text{some function}$,” you could attempt to solve it to produce an exact prediction of the population’s *growth trajectory*, i.e., how N would change through time. The term *deterministic* means that the outcome of such models is completely

determined by the starting conditions and the fixed *parameters* (constants) of the models. For any set of initial conditions, there is one, and only one, result. We will consider the simplest cases of this traditional sort of model, starting with two models of *density-independent growth*, and then adding *density dependence* to produce the celebrated *logistic equation*, in which population size is regulated at a constant “carrying capacity” set by the environment. Later, we will consider two-species variants of these models to treat *interspecific competition* and *predator-prey interactions*. Here’s the critical limitation: These models have to be quite simple to be analytically soluble. In fact, they are invariably *too simple* to be accurate descriptions of how real populations are likely to behave in real-world conditions. They embody simple assumptions that are never fully met in the real world. And if you add additional terms to include more real-world influences, the math quickly gets out of hand. Therefore, these models are only caricatures of reality, but hopefully they capture the essential processes.

The second, more recent, style of population modeling depends on computer-intensive *simulation models*. Here, you do not analytically calculate the one single deterministic outcome. Rather, you write a computer program that encapsulates some rules for population growth, and then run the program to compute a growth trajectory. Because this approach is virtually effortless after the programming is done, one can include more complicated and realistic rules for growth, including complex interdependencies that are impossible to solve analytically. Furthermore, it is easy to make such models *stochastic* rather than deterministic. Stochasticity is the inclusion of *chance variation* in the parameters of the models. In a deterministic model, for example, you would specify a population growth rate that is a constant. That constant growth rate would produce one unique growth trajectory, and that would be your result, plain and simple: a graph of N *versus* time with a single line. But virtually nothing is constant in the biological world! In a stochastic version, you would model the growth rate as a variable rather than a constant. Typically, you would characterize the growth rate as a probability distribution with a specified mean and variance. The computer would randomly select a growth rate from the distribution that you have specified, compute the resulting growth trajectory, then select another growth rate, compute a second trajectory, and keep repeating the process until it had done a large number of runs (or *iterations*), say 1000. Your result would now be this collection of 1000 different trajectories—a graph of N *versus* time with a thousand lines instead of one. Such a graph usually suggests a porcupine having a bad hair day, and is not very simple to interpret directly. But you gain insight by examining the statistical prop-

erties of this *collection of trajectories*. What is the *average* population size reached in the 1000 runs? What is the maximum? Did any of the trajectories result in extinction? If so, what fraction of runs went to extinction? And so forth. This model output is richer and more informative than deterministic output (assuming the model is built in a reasonable way).

Although purists tend to prefer the spare mathematical elegance of analytical models, the advent of cheap and easy computing in the 1980's has made simulation models increasingly popular. Simulation modeling is more like "playing a computer game" than "doing calculus." For many of us, that choice is easy!

Excel spreadsheets available on BIO120 Portal/Blackboard site

Some people can look at an equation and immediately intuit what it implies. Most of us, though, understand the behaviour of equations only after using them to make calculations. I have built some Excel spreadsheets that evaluate the models I present in lecture. I will refer you to particular Excel spreadsheet pages on Blackboard that should help with the concepts I'm trying to explain verbally. If you already know the basics of spreadsheets, these tools should be immediately useful. If you have somehow escaped Excel up to now, I suspect that you can still extract value from this resource. If you have downloaded the ClassApp:Biology, consult it for Excel tips. Otherwise, tutorials are widely available.

Modeling populations without limits: exponential and geometric growth

Here we really start to grapple with the struggle for existence. Charles Darwin, influenced in part by a famous 1798 essay on human population by the English cleric Thomas Malthus, realized that all organisms have the potential reproductive ability to produce multiple offspring. Under ideal conditions, that is, individuals of all species can produce enough descendants to *more than replace themselves*. It is easy to grasp that, given plentiful food and habitat, and in the absence of predators or disease outbreaks, all populations can grow. That critical fact that Malthus and Darwin realized, however, was that growing at a constant, greater-than-replacement rate does not simply lead to population growth. It leads to *explosive* growth.

To make that point dramatically in *The Origin of Species*, Darwin took the slowest-growing species he could think of, the elephant. He estimated lifespans and reproductive rates, and worked through the arithmetic of showing that the descendants of a single pair of elephants would reach staggering numbers in a modest number of generations, assuming that they all kept up a constant positive reproductive rate. Darwin's numeri-

cal answer may not have been exactly right, due to such complications as sex ratios, but his take-home message was correct. Because 21st century university students are more sophisticated than the 19th century laity for whom Darwin was writing, we can demonstrate the principle with some simple mathematical expressions. We adopt some simplifying assumptions to evade such issues as sex ratios.

Geometric growth model

The basic question is how does population size N change over time t , so the equations will all feature those two variables. We will start with the model for *geometric* growth. This model considers time as advancing in discrete steps, such as years, rather than flowing continuously. In this model, for example, the value of t would stay constant at 2017 for a year, then instantly jump to be 2018. In more general notation, we view time as jumping from t to $t+1$ to $t+2$, then from $t+2$ to $t+3$, etc. This model is appropriate for organisms in which reproduction is concentrated in brief annual periods, such as most animals in temperate habitats. (In my examples, I will refer to time intervals as *years*, but this is purely to keep the terminology simple. The intervals could be any time unit that is appropriate to the organism being considered. For single-celled algae, the appropriate interval might be *days*; for bacteria in culture, *hours*, etc.)

As time jumps from one year to the next, how do numbers change? For the simplest case of a closed population with no immigration or emigration, the population size one year in the future (N_{t+1}) will equal the population size now (N_t), plus any births that occur during the interval (B), minus any deaths that occur during the interval (D). (If the population is not closed, we would treat immigration as births and emigration as deaths.) If we just count up numbers of individuals, we can write:

$$N_{t+1} = N_t + B - D$$

However, for modeling purposes, it is better to express birth and deaths as per-capita *rates*. The population-wide birth rate is the average number of offspring that an individual in the population gives birth to in a year. Similarly, the death rate is the average “number of deaths” that an individual suffers per year. These rates will be continuous variables. Obviously, any particular individual can only die once, but if, for example, we expect that half of the individuals in a population will die and the other half will live, the per-capita death rate is 0.5 per individual.

Because these rates embody the number of births or deaths per individual in the population, we can calculate the magnitude of population change by multiplying the rate times the number of individuals:

$$N_{t+1} = N_t + N_t \text{ (per capita birth rate)} - N_t \text{ (per capita death rate)}$$

To simplify this expression, we can calculate a *net* reproductive rate λ by subtracting the death rate from the birth rate. With minor rearrangement, we get the basic model for *geometric growth*, expressed as a *difference equation*:

$$N_{t+1} = N_t \lambda$$

This makes it clear that λ , usually termed the *finite rate of increase* or the *geometric rate of increase*, is *the multiplicative factor by which a population changes from one interval to the next*, i.e., from time t to time $t + 1$. If $\lambda = 1.0$, the population size stays steady. Births and deaths exactly cancel out. If λ is greater than 1.0, the population increases; if it is less than 1.0, the population decreases.

OK, the population increases, but *how* does it increase? Typically, we want to use a mathematical model to make a *projection* into the future from some starting point. By convention, we start at time $t = 0$. At that point, the starting population size is N_0 . For the case of geometric growth, the population grows by a factor of λ each year. As the years accumulate, the λ values pile up multiplicatively. Therefore:

$$\begin{aligned} N_1 &= N_0 \lambda \\ N_2 &= N_1 \lambda = N_0 \lambda \lambda = N_0 \lambda^2 \\ N_3 &= N_0 \lambda \lambda \lambda = N_0 \lambda^3, \text{ etc., or, more generally,} \\ N_t &= N_0 \lambda^t \end{aligned}$$

In the last expression, note how time (t) now appears as an *exponent*. This is showing us that, as each time interval ticks by, the population size gets multiplied by another λ . This means that each interval sees a larger number of individuals added—the population grows faster and faster. For any value of λ *greater than one*, a graph of N versus t for geometric growth is like an ascending stairway, where the riser of each step is taller than the one before it (see lecture slides). If you connect the steps with a smooth line, you get a half-parabola. If λ is small, the parabola is shallow and numbers rise comparatively slowly. Larger values of λ produce faster growth. Obviously, this model describes *unlimited* growth. It can also be described as *exponential* growth in a broad sense, although in population ecology we tend to restrict the term exponential growth to the model described next. Geometric growth is also described as one case of *density-independent* growth. That term emphasizes that the

model includes no crowding effects (density dependence) that would reduce the rate of growth as N gets larger.

If λ is less than one, the population's fate is different. Now, the stairs run downward, and each riser of the staircase is *shorter* than the previous one. The population tapers down toward extinction. Mathematically, the population dwindles to increasingly tiny fractions of individuals, but biologically, it goes extinct when N drops below unity.

Exponential growth model

This characterization is extremely similar to the geometric model. It also describes density-independent growth with fixed birth and survival rates. The only difference is that changes in population size are no longer modeled as annual pulses that produce discontinuous jumps in numbers at each time interval. Instead of the *difference* equation that yields the geometric growth model, here we use a *differential* equation. With that formulation, births and deaths occur *continuously* rather than episodically: population growth trajectories are smooth functions rather than step functions. This entails the adoption of the ideas and notation of calculus. If you are not familiar with calculus, don't panic; for better or worse, in BIO120, we will barely touch its lovely but not-always-welcoming surface.

The distinction between geometric and exponential models can be viewed as a narrowing of the time steps. Imagine a typical description of a species' geometric growth, using years as time intervals. Now, say that you choose to model the same species, but you shrink the time intervals down to days instead of years. Of course, the new λ would be much smaller, but the trajectory would be essentially the same. It would simply have 365 times more steps, and the steps would be correspondingly finer. If you further shrink the time steps down to seconds, the curve gets smoother still. Nevertheless, it would still be a step function if you looked at it with a fine enough resolution. The idea behind calculus is that you continue the shrinkage to the ultimate *limit* where the time slices have zero width. The bumpy staircases become absolutely smooth ramps or roller-coaster tracks.

Suppose you want to describe the speed of population increase at some point along the graph of N versus time. With geometric growth, you would locate the nearest step, and then determine the steepness of that step by dividing its height by its width. You could write that slope as $\Delta N/\Delta t$, indicating a "change in N over a change in t ."

With smooth exponential growth, there is no step, just a curved line. So, to determine the steepness of the curve, you construct a straight line that is *tangent to the curve* at the point of interest. The slope of that tangent straight line is your measure of how steep the curve is, and it is written as dN/dt , indicating an “*instantaneous* change in N over an *instantaneous* change in t .”

Now we can write the exponential model for population growth as a *differential equation*:

$$dN/dt = rN$$

Here, the parameter r is a measure of net growth rate that is closely analogous to (but not the same as) λ . It is often called the *intrinsic rate of natural increase*, or sometimes the *Malthusian parameter*. More casually, ecologists frequently refer to it as “*little r*” to distinguish it from another parameter (“*big R*”) that we have not yet met. You can express r as the difference between the instantaneous birth rate b and the instantaneous death rate d . If the birth rate exactly matches the death rate, then $r = 0$, and the population is steady; that is, $dN/dt = 0$, also (as is obvious from the equation). If r is greater than zero, the population increases, and if it is less than zero, the population declines. Therefore, $r = 0$ is a hinge point for this model, just as $\lambda = 1.0$ is a hinge point for the geometric version.

The advantage of modeling growth in differential equations is that calculus provides numerous solutions to such equations. Because those solutions were developed long before we had computers, they provided the earliest way to work with the models. For example, the differential equation for exponential growth ($dN/dt = rN$) is solved for N versus time as follows:

$$N_t = N_0 e^{rt}$$

The new term e is a constant, the base of the natural logarithms, whom I hope you have already met in other contexts. Like π , it is an irrational and transcendental number that goes on forever when expressed as a decimal fraction. Its value to three places is 2.718, but all you need to know to interpret the above equation is that e is a constant. The *critical* thing is that t now appears as an *exponent*, as it does in the geometric case. This again means that a growing population grows with ever-increasing speed in a “population explosion,” whereas a declining population dwindle toward zero in an exponential decay.

Thought/calculation question. Rates of exponential growth are frequently characterized by calculating the “doubling time” of population size, and rates of exponential decline are frequently characterized by calculating the “half-life,” or the amount of time it takes for the population size to decline by 50%. How would you calculate these numbers?

The logistic model of density dependence

I'll spend time in lecture trying to reinforce the ironic conclusion that real populations do not long behave according to either the geometric or exponential models, with fixed values of λ or r . If those parameters stay below their hinge points for long, species go extinct; if they stay high, unending population explosions would cover the earth. We do see some extinctions, but population explosions are rare. When they do start to occur, they always get stopped eventually by some aspect of crowding, such as starvation, lack of shelter, or disease. There is, therefore, a tendency in nature to *regulate* the extremes of population growth. Actual mechanisms of regulation are numerous and complex, but population ecologists have long used simple *density-dependent* equations to portray mathematically simple models of regulation. By far the most widely used is the *logistic* model, which is commonly written as a differential equation in one of two equivalent forms:

$$\frac{dN}{dt} = rN\left(\frac{K - N}{K}\right) \quad \frac{dN}{dt} = rN(1 - (N/K))$$

I strongly suggest that you familiarize yourself with this equation by plugging in various values for its three parameters and seeing how the growth rate responds. The BIO120 Blackboard/Portal site offer an Excel sheet that makes this easy.

This iconic equation is only one of several ways that one could model density dependence, but it has become entrenched in the theoretician's toolkit. As we will see, it serves as a foundation for more complex models. Its main virtue is mathematical simplicity rather than realism, but it is a huge step more realistic than exponential growth! How does it work? It is obviously based on the exponential growth model. Note that one way to model density-dependent crowding would be to change r from a constant to a variable, and to make r get smaller as N gets bigger. That is not how the logistic model works. Instead, we add a new term to the model that acts to put the brakes on dN/dt while leaving r as a constant. The new term introduces a new parameter K , which is termed the *carrying capacity of the environment*. K is defined as the *number*

of individuals that the local environment can support indefinitely at a steady state. I think the best way to envision this concept is to imagine that the resources needed for growth and reproduction sort of ooze out of the environment at a constant rate. They become available to the organisms in our population at this constant rate. When N is very small, there are plenty of resources available for each individual, and those individuals reproduce at a very high rate. The population grows *almost* exponentially. As N gets close to K , however, most of the resources are being used up, and growth slows to a crawl. As I will show with graphs in lecture, the population size N approaches K asymptotically. Whether it is growing from below K or declining from above K , N approaches K and stays there. The solution of the differential-equation form of the logistic is more complicated:

$$N_t = \frac{KN_0e^{rt}}{K + N_0(e^{rt} - 1)}$$

You do not need to memorize this rather prickly expression, but *you do need to know the qualitative shape of the graphs of N versus time* that it produces. I will present that material in lecture, but the best way to explore this equation (if you have the skills) is to set it up as a formula in a spreadsheet or on a graphing calculator. If you lack the skills to set this up, you can consult the ready-made Excel spreadsheets on the BIO120 Portal site. (Again, if you are mystified by Excel, consult the aforementioned ClassApp:Biology app or find some online tutorials. Now is as good a time as any to come to grips with some spreadsheet program. You can't avoid spreadsheets if you stay in science.)

By entering various starting values of N_0 , and playing around with the constants r and K , you can see how they affect trajectories for N . To ruin the suspense (spoiler alert!), if N starts low enough, the graph of N versus time has a classic sigmoidal (S-shaped) form: numbers initially rise nearly exponentially, then slow down. The trajectory hits an inflection point at $N = K/2$, and then decelerates toward K .

In various textbooks, you may read offhand statements that the logistic equation *always* generates a sigmoidal approach to K , but that's not quite the whole story (see **Figure 7**, page 6). If N starts only a modest amount above or below K , its trajectory for approaching K will lack an obvious inflection point that characterizes a sigmoid curve. Check this out for yourself! What we can say, however, is that a population growing logistically will always make a direct and smooth approach to the carrying capacity, from any starting point other than $N = 0$.

This is a radically different view of the world than we got from the exponential or geometric models! Those models suggested great instability in population trajectories, with species overfilling the globe or crashing into oblivion. The logistic suggests a much more serene world, with each species comfortably settling into its carrying capacity, breathing a sigh of relief, and staying there in a permanent equilibrium. This kind of model probably has contributed to the rather pervasive belief that there is a strong *balance of nature*, and that density-dependent processes keep natural communities basically *equilibrium* and *stable* in the absence of extrinsic destabilizing forces (such as malign human influence). It's true that there are some *regulatory processes* at work, especially crowding effects, but the idea of a serene and balanced state of nature can mostly be dismissed as New Age Crap. As I'll discuss in the upcoming section on community ecology, an equilibrium viewpoint was common in the 1970's, but ecologists have increasingly been turning away from that expectation.

One reason for this change of tone is the accumulation of case studies in which the logistic equation is a very poor model of reality. In the early days of population ecology (the 1920s), the elegant mathematics of the logistic convinced some influential researchers that it *ought* to be a law of nature, rather than a quantitative hypothesis about how density dependence might possibly work.

That impression was bolstered by a number of early empirical studies on organisms like protozoa, growing in culture vessels in the lab, with a steady input of food. Those culture conditions frequently were good matches for the implicit assumptions of the logistic model. Unsurprisingly, growth trajectories often were sigmoid, as populations grew from small to crowded. However, investigators tended to terminate their experiments as soon as population growth seemed to level off. (Scientists in all fields often make the mistake of stopping their observations as soon as they have seen what they were expecting to see.) If these cultures of protozoa had been continued longer, they would frequently undergo population crashes due to factors such as the buildup of toxic wastes. If the logistic equation were a truly robust and reliable description of how populations always behave, it should be rather easy to maintain a culture of yeasts or *Paramecium* at a steady value. In fact, it is quite hard, usually requiring a special flow-through vessel called a chemostat.

Figure 8 (page 6) by Pearl and Gould gives an almost humorous example of early ecologists' peculiar faith in the logistic model as a natural law. It's cited in the first major textbook on animal ecology (Allee, Em-

erson, Park, Park, and Schmidt, 1949). The authors gathered all of the existing estimates they could find on the human population of the world through time, and used statistical methods to compute the logistic growth curve that would best fit the data. They calculated that the number of people would level off at a K value of about 2.6 billion. The actual number recently passed 7 billion and is still climbing fast. If you are reading from a hard copy, I suggest that you use a red marker to extend the scale of the graph above to include a data point for $t = 2014$ and $N = 7$ billion. (*Is your paper big enough?*) Obviously, the logistic was a laughably poor model of human growth over the last several decades.

Thought questions: “Curve fitting” is a universal process in scientific research. You will encounter it throughout your career, but it has notorious pitfalls. What made Pearl and Gould’s prediction go so wrong? Consider both biological and statistical aspects, insofar as you can. What are the broader lessons concerning the risks of fitting models to data? Note that the UN is now offering models that indicate global population leveling off at about 9 billion. Are those predictions more believable now than the above graph was in 1949?

Gratuitous advice: Science majors at U of T need to take some stats, typically in year 2. Some students fearfully postpone this course until later. That’s a terrible idea—don’t do it.

What about populations in the wild? All species show variability in numbers through time. Among the most stable populations are relatively long-lived, large-bodied animals with low reproductive rates. For example, tropical bird species that exhibit territorial behaviour in habitats with stable weather would be expected to keep rather constant numbers. If all the good territories with nesting sites are taken up by territorial adults, a new juvenile member of the population may not be able to breed until a territory holder dies. At the other extreme, insects are notorious for having population outbreaks, or *irruptions*. As we have discussed, small animals are usually obligate ectotherms, and are poorly buffered against adverse temperatures. But most insects have very short generation times, giving them very high potential r value if the weather is favourable and their food supply is plentiful. The classic case is the occasional outbreaks of migratory locusts in Africa, where such unbelievably large numbers of insects are produced that they completely wipe out all vegetation. Similar—but less dramatic—effects happen in North America with *forest pests* like the spruce budworm or the gypsy moth, which wipe out many trees during irruptions, and many *agricultural pests* like the corn earworm caterpillar.

These outbreaks clearly represent population dynamics that are not well described by the logistic equation. We are *not* seeing gradual, inevitable approaches to equilibrium. That does not mean, however, that there is no density dependence at work. In fact, violent population outbreaks of insects are usually followed by violent population crashes. Sometimes the insects simply consume all the available food, and then starve. In the case of gypsy moths, caterpillars can be infected by a polyhedrosis virus. A sick caterpillar fills up with virus and then disintegrates or explodes, spreading virus particles into the habitat. The spread of infective particles through the environment is enhanced by the tendency of the virus to turn caterpillars into wandering “zombies” that climb to the treetops before expiring and exploding. As viral particles becomes ubiquitous in the environment, few caterpillars escape infection. Yes, the disease “goes viral.” This very thorough dissemination of disease in crowded conditions helps terminate an outbreak. Worsening weather usually helps, too. By simple statistical logic, a period of *unusually good weather* for an outbreak is likely to be followed by more normal weather, which will temper the population growth. Returning to our dichotomy of *density-dependent* versus *density-independent* effects on population, we would consider the weather to be *d-i* and the viral disease to be *d-d*.

As you might expect, there are detailed mathematical models for the spread of disease, especially for important human diseases. Density dependence of transmission is the key factor. These models get very interesting (and complicated) for diseases in which some individuals survive infection and become resistant. Variation in transmission probabilities carries important implications for why some diseases are mostly benign (the common cold virus) while others are devastatingly virulent (Ebola virus). BIO220 will take up some of these topics, including the implications for medical practice.

More flexible forms of density dependence

The kind of density dependence we see in erupting insects is clearly not logistic. But can we write alternative models that capture a more realistic range of dynamics? Of course, the answer is yes. Somewhat surprisingly, such models do not need to be more complicated than the logistic; that is, they do not need any more parameters. Here is a celebrated *difference-equation model popularized by Robert May*, a theoretical ecologist who has also contributed much to disease modeling. It starts with geometric growth, but adds a carrying capacity to which the population returns by density dependence. So far, everything is analogous to the logistic. The key innovation in this model is that the difference-equation framework intrinsically allows *time delays* in response. In the standard differential-equation logistic, the density dependence is applied imme-

diately: at every instant in time, population growth is adjusted infinitesimally, depending on whether the population is above or below the carrying capacity. Therefore, growth shows *perfect compensation* for the difference between N and K , and the population size sneaks up on the carrying capacity gradually (technically, asymptotically). In the difference-equation formulation by May, the population size N is compared to K at a particular time interval. If the population size is well below carrying capacity, there will be compensating surge of growth. But the added individuals don't show up in the population until one time interval later. There is no guarantee that this increment of growth brings the population exactly to K . It might reach to less than K , in which case a gradual, logistic-like approach to K is possible. But it might also exceed, or *overshoot*, the carrying capacity. In that case, we will see new dynamics. The equation is:

$$N_{t+1} = N_t e^{(r(1-N_t/K))}$$

Let's break this expression down. As before, e , r , and K are constants. The whole expression $e^{(r(1-N_t/K))}$ takes the place of λ as the *factor by which population size changes* over the time interval we are examining—but unlike λ , it is not a constant. Rather, it varies depending on the relationship between the current population size and the carrying capacity. The $(1-N_t/K)$ term in the exponent is the same as $(K-N_t)/K$, which you should recognize as being the same as the “braking term” that we discussed for the logistic equation. It indicates how far the current population size is from the carrying capacity. It also indicates, by its sign, whether the current population size is above or below the carrying capacity. This term is multiplied by r , which we previously saw as a measure of a population's ability to grow. When it appears in this equation, its role is better understood as measuring the *strength of the density dependence*. Bigger values of r mean that populations will rebound more strongly toward the carrying capacity, whether they are above it or below it. Therefore, bigger values of r are more likely to produce *overshoots* and complicated dynamics. Because overshoots drive the population above K , they will be followed by undershoots.

To see this graphically, set up a spreadsheet to examine the behaviour of May's model (or, less desirably, use the one provided through the BIO120 Portal). If you start with low values of $r < 1$, you will first see smooth approaches to carrying capacity, much like the logistic. As r becomes greater than one, however, you will start seeing *damped oscillations* around the carrying capacity. At this intensity of density dependence,

the swings in population size resemble the swings of a pendulum that are gradually diminished by friction: the system eventually comes to rest with $N = K$. If you continue to dial up the r parameter, you will see the damping disappear. At first, you will see *stable limit cycles*, in which oscillations are not damped, but rather continue forever like the swings of a *frictionless* pendulum. When r reaches values of 4 or so, the repeatability of the cycling breaks down entirely. The system enters the realm of mathematical *chaos*, in which fluctuations become violent and inherently unpredictable. Like terrifying roller coasters, numbers soar to great heights and then swoop down to nearly zero. Extended chaotic trajectories have properties that greatly interest mathematicians (see Wikipedia for cool animations!), but the ecological implications are mundane: a chaotic trajectory will quickly send a population to extinction because numbers will reach low points of less than one individual. The take-home message is that *weak* density dependence leads to smooth, logistic-like behaviour, but that *stronger* density dependence leads to increasingly violent overshoots. The same effects can be obtained with different models that explicitly incorporate *time lags* in response to deviations from carrying capacity. Rapid responses give smooth reactions, whereas longer time lags in the action of density dependence produce overshoots. May's difference-equation formulation doesn't include an explicit term for time delay, but can still be interpreted as implicitly incorporating time delays: as r increases, *more population change happens* per time interval.

Thought question: Go back to the graph from Allee et al., (Figure 8, page 6) where I hope you extended the population trajectory up to 7 billion. Can we regard that trajectory as a population overshoot of the type predicted by the May model? Why or why not?

Inverse density dependence: Allee effects

So far, our models for density dependence have been concerned with the negative consequences of *high* numbers. May's equation and the classic logistic both produce maximum individual growth rates when the population is at its lowest numbers. Any increase in numbers above zero exerts a crowding effect that hurts individual performance. Is this realistic? Possibly it might be for populations such as yeast or bacteria, but most higher animals and plants will usually suffer if their numbers fall too low, also. The most obvious downside of low density lies in finding mates. To consider an extreme example, imagine that you are the only member of a population; $N = 1$. Well, you are certainly not crowded, and there would be plenty of food resources for any offspring you might produce. The logistic would tell you that *per-capita growth rate* should be

at a maximum. However, you are not going to produce any offspring because you don't have anybody to mate with! Even if we relax the scenario and imagine a population of more than one, you may still have trouble finding a mate if your conspecifics are spread too thinly across a large landscape.

Animals are typically mobile, and often very good at locating mates through exhaustive searching and remote detection, so the problem is less severe for them. Imagine now that you are a member of a small and sparse *plant* population. Your mating success probably now depends on getting a visit from a pollinating animal. Pollinators are unlikely to specialize on visiting plant species that are few and far between, so even if you get a visit, the pollinator may not be carrying pollen grains of your species. If you are adapted for pollination by wind, your situation is almost as dire.

W.C. Allee, the prominent early ecologist who cited the graph you encountered above, was especially interested in the general class of *positive* social interactions in which members of populations facilitated each other's survival and reproduction, instead of harming each other through intraspecific competition. He reasoned that negative effects would inevitably predominate at high densities, but that positive effects might prevail when N was low. Mate-finding was perhaps the most common and general, but he compiled other examples, too. Animals often congregate in groups (flocks of birds, schools of fish) that confer mutual protection against predators. Protection against predators can occur through warning alarm calls, through dense schooling behaviours that confuse predators and deflect attacks, or through ganging up to offer a stronger defense against attacks. Many animals pick up cues about food locations from conspecifics' behaviour. By flying in a V-formation, geese can "draft" behind others and fly with less energetic cost. In some cases, animals work together to provide protection against physical stresses. Allee did experiments showing that fish in aquaria could withstand higher concentrations of poisonous metal ions if the accumulated toxic load could be spread across more bodies. Groups of fish could survive in solutions that would kill a solo individual. Collectively, all of these diverse responses are called "Allee effects." (Small populations are also susceptible to the negative genetic effects of inbreeding, but this topic is treated in the evolution portion of BIO120.)

Allee effects are of special importance in conservation biology, because conservation problems are almost entirely concerned with rare species. If the effects are strong enough, there can essentially be a *lower threshold*

of numbers, below which the population declines instead of growing. You can modify the logistic by adding a second “braking term” as follows:

$$\frac{dN}{dt} = rN \left(\frac{N}{A} - 1 \right) \left(1 - \frac{N}{K} \right),$$

Here, the constant A defines the *Allee threshold* just as K defines the carrying capacity. If the population size N dips below A , dN/dt goes negative, and the population will decline to extinction. Do not concern yourself with the solution to this more complicated expression, but do consult lecture slides for illustrated trajectories. In practice, it would be very hard to determine an exact value of A for any specific population of an endangered organism. And if you *did* make an estimate of A , it would be very risky to trust your estimate by letting numbers decline toward that level. Therefore, calculating a numerical value for A would probably be a sterile exercise in any real program of conservation. Nevertheless, it would be essential to understand the biology well enough to know what sort of Allee effects would likely be important.

Heuristic models versus management tools

It is important to realize that the fluctuations and oscillations in these models arise entirely from density-dependent processes intrinsic to the population; the environment is held constant, so there is no weather or seasonality involved. The equations aim to capture the “something” that causes density dependence, but they accomplish this through the abstraction of the “carrying capacity.” There is no explicit separation or identification of possible limiting factors such as food supply, space to live, or susceptibility to disease. In the real world, things are more complicated. Populations are certainly buffeted by bad weather and buoyed by good weather. Seasonality affects individual survival directly through temperature or osmotic stress, and indirectly through food supplies. For most animals in temperate biomes, the carrying capacity is probably set by winter conditions, not summer.

For these reasons, the simple models we have looked at would never be very useful for practical applications like predicting or controlling outbreaks of an agricultural pest species, or conserving an endangered species. For such purposes, you would need to arduously construct a *species-specific* model that would include many more parameters about reproductive rates, weather, food supply, and other vulnerabilities. To build such a model is to wallow deeply and endlessly in a swamp of picky details. Such enterprises are very instructive, but not appropriate for an introductory course like BIO120. We will stick with simple mod-

els of *heuristic* value (i.e., “involving or serving as an aid to learning”), rather than practical management tools.

We have now covered the major points of simple, single-population, deterministic growth models. From this point, we will introduce added sophistication in three ways: we’ll add *stochasticity*, *age structure*, and *multiple species*. There is no obvious logical ordering to these topics; following the structure of more advanced EEB courses, I will delay considering multiple species until we formally move from population ecology to community ecology. I’ll start with age structure, because it provides an opportunity to enliven our ecological focus with some fresh perspectives on the evolution of life histories.

Age-structured models of population growth

So far, our models have considered a population of N individuals, all of whom are identical with respect to their prospects for survival and reproduction. This keeps the math simple, but of course, all individuals are *not* identical in real populations. From the standpoint of population growth, the most important differences among individuals are *sex* and *age*. We know that the probabilities of dying and of giving birth are not the same for all classes. In age-structured models of population growth, we treat age differences by dividing the population into a convenient number of different *age classes*, each of which will have *different, age-specific* prospects of death and reproduction. The formulas that follow amount to a bookkeeping exercise to keep track of the different age classes within a population. Age-structured models deal with sex in a simpler way: they keep track of the females only, utterly ignoring the males. This isn’t as arbitrarily dismissive as it sounds. Females are the only ones who reproduce, so they are the only propulsive force for population maintenance. To understand things like population growth rates and generation times, it is pretty much acceptable to regard males as a numerically inconsequential collective source of sperm—a source that is unlikely to be in short supply. And since many animals have sex ratios that are close to 1:1, it’s possible to do the math on females and just multiply by 2 to estimate the total population of both sexes. Therefore, we will build models in which females produce daughters. Males simply don’t appear in this bookkeeping. The population size will be steady if the average female produces one daughter during her lifetime.

Life tables: math and notation

As previously mentioned, we begin by assigning the individuals in a population to classes based on their age, using the subscript x to stand for age. The time span of an age class interval is chosen for convenience: we want enough age classes to capture the way that birth and death prospects vary through a female’s life, but too many classes will make cal-

culations burdensome. In life tables for humans, each age class is typically chosen to be 5 years long, so that about 20 age classes are enough to capture the human lifespan. For organisms with shorter life spans, one could use years, months, weeks, or whatever interval fits the biology. To simplify my verbal descriptions, I'll describe sample calculations as if age classes are years. By convention, the first age class is denoted by the subscript 0. Thus, the total population of N is divided into n_0 newborns, n_1 one-year-olds, n_2 two-year-olds, etc. If we sum up the n_x values over all ages x , we get the total population size. In summation notation,

$$\sum_x n_x = N.$$

For this and all the following cases, the capital sigma indicates that you sum up the following variable over all possible age classes. The set (or *vector*) of n_x values (that is, n_0 , n_1 , n_2 , etc.) is called the *age structure* of the population at the time we are looking. It is traditional to graphically represent the age structure as a horizontal bar graph called an *age pyramid* (see lecture slides).

In our previous modeling of population growth, we looked at the net balance between births and deaths. We will proceed to do the same here, but we can no longer assume that birth and death rates are uniform constants for the whole population. We now have different rates for each age class.

Survivorship schedules

Now, we summarize the age-specific risks of mortality as the *survivorship schedule*, by defining l_x = survivorship at age x = the probability that an individual is still alive at age x . The use of l_x , with a lower-case letter l , is a convention that I will follow in this text. In lecture slides that use the sans-serif Arial font, I may use an upper-case L to avoid confusion with the numeral 1.

Some things are necessarily true about l_x values. Because they are probabilities, they range from 0 to 1. Because newborns are by definition alive at birth, $l_0 = 1$. And l_x values must always get smaller as x increases, *i.e.*, the probability of being alive always declines with age. Ultimately, everybody dies. For a life table to be useful, it has to include enough age classes for the last l_x value to be zero. The shape of the survivorship schedule, *i.e.*, a graph of l_x versus x , is an important component of an organism's life history. Every species will have its own characteristic shape, as I will discuss in lecture. Some species produce huge numbers of tiny offspring that have few defenses against environmental stresses. Those undefended weaklings will mostly die early in life, giving a survivorship curve that initially drops steeply. The few that survive, how-

ever, can grow large enough that life becomes less hazardous, and survivorship levels off. Other species produce fewer offspring, but invest more resources in each of them. These larger offspring have the resources and resiliency to withstand environmental stresses, which will be reflected in a survivorship curve that initially descends more gradually. No organism lives forever, though, so these too will eventually die off, not so much from environmental stresses as from the generalized breakdown of old organisms that we call *senescence*.

We can mathematically specify a particular type of survivorship curve that can serve as a reference point for comparing different species. Imagine a hypothetical species for which the probability of dying is constant across all ages. This is tantamount to death coming by external accidents, rather than from intrinsic biological characteristics such as infant mortality or senescence. This type of survivorship would characterize processes like the radioactive decay of unstable isotopes, or the expected lifespan of wineglasses in a restaurant. Mathematically, such distributions fit an *exponential decline* with a constant half-life. A graph of l_x versus x gives a curve that initially declines steeply but then gradually approaches zero as an asymptote (just like the extinction of an exponential population with $r < 0$). Traditionally, such distributions are called Type II survivorship curves—not a usefully descriptive name, but seemingly entrenched. For a differently scaled visual display of an exponential Type II decline, we can take the logarithm of survivorship. If you plot $\log l_x$ versus x , you get a straight line with a negative slope. The convenience of this property has led to a convention of plotting survivorship curves on logarithmic scales. If this is done, we call straight lines Type II, concave curves (with much early death) Type III, and convex curves (low early mortality) Type I. (Note: I include this comparison of linear plots versus log plots *not* because it is essential to understand how life tables work, but because this issue continually pops up in both the physical and biological sciences. It's good to practice thinking about it.)

Fecundity schedules

Having defined a way to handle age-specific probabilities of dying, we now use a similar structure to quantify the expected amount of reproduction. We define b_x as the average number of daughters produced by a female in her x th year of life. (Note: this quantity is often symbolized by m instead of b , especially in older literature. I may not have purged the m notation from all of my lecture slides.) Because b_x values are not probabilities, they can exceed 1.0. Typically, b_0 will be 0, unless the age-class intervals have been chosen to be inappropriately wide; newborn organisms need to pass through a period of resource acquisition and

growth after birth before they are capable of giving birth themselves. As for survivorship, the schedule of fecundity will differ for different organisms. The key elements are the total number of offspring produced and the waiting period before reproduction can begin, typically termed the *age of first reproduction*. Importantly, in these simple age-structured models, we will treat both fecundity and survivorship schedules as constants. This means that these models *lack* density dependence.

To calculate population-wide rates of increase, one must consider how fecundity and survivorship interact with each other. We do this by considering the expected lifetime reproduction of the *average female*. It's critical to understand this concept, so work through the following argument until it makes sense. First, imagine a female who gets lucky, and lives the longest life possible. The number of daughters she will produce in her lifetime is simply $\sum b_x$, the sum of all of the age-specific fecundity values. But not all females are so lucky. One who dies before she reaches the age of first reproduction will produce no daughters at all. One who dies in the middle of her potential lifespan will produce some intermediate number of daughters, less than $\sum b_x$ but greater than zero. To calculate the lifetime output of the *average female*, we need to downweight or "discount" her potential reproduction at a particular age by the likelihood that she dies before reaching that age. We accomplish that discounting by multiplying each b_x value by the corresponding l_x value, then summing up those products across all age classes x . That gives us the expected number of daughters a female will produce in her lifetime, a quantity usually called the *net reproductive rate* or the *net replacement rate*. It is symbolized as R_0 . The formula, which you must remember, is:

$$R_0 = \sum l_x b_x$$

Now, you should recognize that R_0 can act similarly to the finite rate of increase λ . Both are multiplicative factors by which the population grows from one point in time to the next. If R_0 is equal to 1.0, each female exactly replaces herself, and the population size will remain constant. If R_0 is greater than 1.0, the population will increase geometrically without limit. If it is less than 1.0, the population slides toward extinction. How R_0 and λ differ is in the time scale over which they measure growth. As you should remember, λ is defined as the amount of growth that occurs over one *unit of time*, such as a year. (Recall that the unit of time is arbitrary; we could choose months or days and get a different λ .) But R_0 is the amount of growth that occurs over one *generation*. This should make some logical sense, because it is equal to the number of daughters that a typical female will produce to replace herself. How long does that replacement take? Approximately a generation.

We can get rid of that “approximately” if we define the generation time mathematically to be the *age of a mother at the time she produces her average daughter*. This requires computing a weighted average, which can be tricky if you are not familiar with the concept. You can just memorize the following formula, but it will be more useful to you to work through the logic—you will encounter weighted averages frequently in a scientific career. Suppose that the typical mother produces a single daughter when the mother is two years old. In that simple case, the generation time is 2 years. But suppose that a female can produce one daughter when she is two and two more daughters when she is three. Clearly, the generation time is going to be greater than 2. It might be tempting to call the generation time 2.67 years—the mean of $2 + 3 + 3$ —but that leaves something out: some mothers will die between years two and three, so, on average, more than 1/3 of the offspring will be produced by two-year-old mothers. This means that offspring production by older mothers must be downweighted by their lower survivorship. The formula for generation time T reduces to:

$$T = \Sigma x l_x b_x / \Sigma l_x b_x = \Sigma x l_x b_x / R_0$$

Note that the first x after the summation sign in the numerator is *not a subscript*, although the other x ’s are. *That means that the first one is a value, not just a label.* Consult the “Life table 1 parameters” tab on the “Population modeling spreadsheet” on the Portal to verify that you can work through the calculations.

Projecting population growth and age structure from life-table parameters

If you know the numbers of females in each class at some starting time t , you can model the future growth of the population under the assumption that the fecundity and survivorship values are fixed. An additional assumption is required to simplify the calculations for discrete time units: all the deaths and births that occur during a time interval take place at the *last possible instant before the next interval begins*. If several cohorts of mothers enter a time interval, they all survive to the end of the interval—say, one second before midnight on the last night of the year. They then all have babies according to the b_x schedule. At midnight, some fraction of the mothers die, according to the l_x schedule. The surviving mothers and all of the new babies register in the next time interval. And the surviving mothers are now one age class older. The babies are all the same age, of course, $x = 0$. This makes the population grow or shrink as a step function, just as we saw with geometric growth.

With these assumptions, you can crank through the arithmetic by brute force: it is simple bookkeeping. Moms die, babies get born, everybody

gets older. But if you know any linear algebra, the necessary calculations can be accomplished through matrix multiplication. The famous “Leslie matrix” is a square matrix in which the top row holds the b_x values, and subdiagonal holds survival data in the form of s_x values, where $s_x = l_{x+1}/l_x$. Postmultiplying this matrix by the vector of the current age structure at one time t will give the age structure at the next time interval $t + 1$. This process can be iterated indefinitely, thereby projecting future changes in population size *and* in age structure. An example in Excel is shown on the “Life table 2 parameters” tab on the “Population modeling spreadsheet” that is available on the Portal. I won’t ask you about the linear algebra approach, but I do ask you to understand the fundamental behaviours of such models. The best way to gain this understanding is to use the “Life table 2 parameters” spreadsheet. First notice how the spreadsheet calculates the projection of the age structure, and also graphs it. (It also calculates the summary parameters λ , R_0 , and the generation time T , which should be familiar.) If you change the fecundity values, survivorship values, or starting population structure, the output displays will change accordingly. Here are some things to investigate:

- Change the *initial age structure*, and see how the projection changes. Also watch the values of λ , R_0 , and T . Do you expect them to change, also? Why or why not? If you are having a hard time answering this question, go back to examine the formulas for λ , R_0 , and T . *What do those values depend on?* Read the comment blocks on the spreadsheet to further investigate the effects of *initial age structure*.
- Compare the projection of *numbers* (blue cells) to the projection of *proportions* or relative numbers (lavender cells). Be sure you understand that, to follow a cohort of individuals through life, you go *down the diagonal* of the table of numbers. With the passing of each year, time increases and age increases.
- Using the default age structure (starting with $n_0 = 100$ newborns), note how the actual increase in size from one time to the next (“lambda,” the bottom row of the numbers projection) is initially bumpy but later smooths out. Note also how all of the age classes become represented, and how their relative proportions become stable. These two phenomena are linked. The age structure converges to a *stable age distribution* (s.a.d.) in which the proportion of the population in each class stabilizes. And the value of λ stabilizes when this convergence is achieved. This convergence happens with any Leslie matrix, as long as births occur in more than one age-class interval. (You can check this out by changing all but one b value to zero, and looking at the projections.) At s.a.d. in a growing population, each age class is increasing exponentially, and all of them are increasing at the same rate. If any l_x or b_x values are changed,

the s.a.d. will also change, and the population will re-equilibrate at the new values.

- Try changing the fecundity and survivorship schedules to achieve $R_0 = 0$, or “zero population growth.” How does the stable age distribution shift with changes in the population growth rate? How do these results link to the age pyramid graphs for the Scottish human population, shown in lecture slides?

A few aspects of life history evolution

The field of *evolutionary ecology* considers, among many other issues, the ecological circumstances that drive selection for such *life history characteristics* as lifespan, the age of first reproduction, and the number and timing of offspring. I will treat some of these issues in lecture because I want you to understand that ecology drives adaptive evolution not just in physiological and morphological characters, but also in *timing characters*; here, I give some notes on that material. If these considerations interest you, BIO220 will go into greater depth with regard to both life-history theory and its applications to humans. Higher-level EEB courses go even further.

Key tradeoffs

Life-history responses to environmental challenges are subject to constraints and tradeoffs. Most of these have something to do with the allocation of limited resources. For example, the production of offspring is affected by a *size-number tradeoff*. Consider a plant with some fixed amount of energy and nutrients that can be invested in seeds. It can either make many small seeds or a few large seeds. What should the plant do? How do ecological circumstances determine the best (or, more precisely, the *optimal*) strategy?

Another tradeoff involves *early versus late reproduction*. Newborn organisms undergo a pre-reproductive period of growth and resource acquisition before they can produce offspring. All else being equal, early reproduction would always be favoured by natural selection, because individual organisms that produce offspring earlier than others will come to dominate a population. But that key proviso “all else being equal”—you may also encounter the Latin equivalent, *ceteris paribus*—seldom applies in the realm of life history. In this case, organisms that reproduce earlier do not have as long a period to accumulate resources, so they can’t make as many offspring as those others who wait longer to start. Again, we expect natural selection to produce a waiting period that is optimal in the current environment.

A third basic consideration is usually termed the *cost of reproduction*. Producing offspring does not come free for parents. Those offspring re-

quire energy and nutrients that the adult could use for other purposes, including extending its own lifespan or saving to make other offspring in future. In many higher animals, the costs of reproduction extend well beyond the birth of the offspring, in the form of *parental care*. In mammals, for example, lactation for feeding offspring often requires more maternal resources than the pregnancy itself! So, producing offspring is costly to parents in two ways. First, parental survival can be reduced. This is a case of the b_x schedule affecting the l_x schedule. Second, parental resources for reproduction can be depleted, meaning that subsequent offspring production is delayed and diminished. In this case, high b_x early in life reduces b_x late in life. Both are classic trade-offs: you can't have everything.

Reproductive value

The great evolutionist, geneticist, and statistician R.A. Fisher introduced the concept of *reproductive value* as another aspect of life history that is a function of the survivorship and fecundity schedules. Reproductive value v_x is defined as the *expected number of daughters to be produced by a female of age x , now and for the rest of her lifetime*. In the previous sentence, “now” means during her x th year of life. Reproductive value is a statement of the current value of an organism for future fitness. For example, if you wanted to buy a thoroughbred mare for the purpose of producing foals, you would want your investment to purchase an animal with a high current reproductive value. Any offspring produced *before* age x are irrelevant to v_x . The concept of reproductive value only looks forward in time. I won’t present a formula because it requires a more complicated summation notation than we have been using, but I do present the calculations in the “Life table” tab on the “Population modeling spreadsheet” that is available on the Portal. Essentially, you calculate v_x by rescaling the l_x values *as if the organisms were starting life* at the age for which you are calculating the reproductive value. Following through those calculations should help you understand this somewhat tricky concept.

Quick thought question to see if you have been paying attention: We have previously discussed the quantity v_0 under a different name. What is it?

Why do I consider v_x a tricky concept? Because a graph of v_x versus x typically gives a humped shape, reaching a maximum somewhere between youth and old age. Part of that humped shape is easy to understand: It is intuitively obvious that reproductive value would decline late in life. The organism’s vigour is declining, and there isn’t much time left for reproduction. But why should a newborn, with *all* her life

ahead of her, have a lower future value than a somewhat older female, who has only *part* of her life ahead of her? The reason is that a newborn must get through the pre-reproductive, resource-acquisition phase of her life before she can start reproducing, and she encounters some risk of dying during that period. Reproductive value, like R_0 , concerns the average female, so the unfortunate ones who die must be averaged in with the others. That is why the v_x of a newborn is lower than that of a female who has successfully passed through the gauntlet of early-life risks. Try changing early-life survival values in the age-structure spreadsheet, and watch how the shape of the v_x curve responds.

One might want to compare the shapes of v_x curves of different species or populations. Shape comparisons are made easier by converting raw values to *relative reproductive values*. This “relativization” is accomplished by dividing all of the v_x by v_0 . The relative reproductive value of all newborns is thereby scaled to equal 1.0. This rescaling facilitates comparing the *shapes* of different curves.

Why do we care about reproductive values?

This concept is most important in population genetics, where variation in v_x among individuals essentially predicts their genetic contributions to future generations. We will not go into the details of such models, but it should be clear that selection can favour age-dependent behaviours. Some important ecological/behavioural aspects include mate selection and the effectiveness of colonists. In species with a tendency to form lasting *pair bonds* between mates, there should be fairly strong selection for males to prefer female mates with high reproductive values. This could produce a tendency for traits associated with the onset of reproductive age to become associated with sexual attractiveness. Human behaviours are so heavily influenced by cultural influences that it is always controversial to invoke genetic causes, but it seems reasonable to ascribe the persistently youth-obsessed patterns of human mate preferences to underlying variation in reproductive value, in part.

Many mobile animals characteristically undergo a temporary life-history stage of migration. Typically, this occurs in juvenile animals near the onset of reproduction. Juveniles are frequently driven out of their natal territories by disputes with their parents, and have to strike out to find territories of their own. It makes sense that this restless period would coincide with a high v_x value, because such individuals would have the best chance of establishing a successful new colony and lineage. Such effects are somewhat weakened in species where males are more likely to be the ones that disperse, as is commonly the case. The same principle is especially important in cases where captive breeding programs are

being used to help the conservation of endangered species. If one is releasing captive-raised animals into nature to start or to augment wild populations, choosing individuals with a high v_x will give the best chance of success.

The big bang theory— ecological circumstances favouring semelparity

In both plants and animals, some species are genetically programmed to reproduce only once in their lifetimes and then die (*semelparity*), whereas others potentially reproduce numerous times (*iteroparity*). In plants, different terminology is used. Iteroparous plants are usually called *perennials*. Semelparous plants get various names depending on how long they live before flowering and dying. All of them are called *monocarpic* (making fruits once). If they typically live only a single season, they are called *annuals*. Other species are programmed to grow vegetatively for one year, and then flower and die in their second year. These are called *biennials*. (In stressful conditions, biennials frequently take some extra years of growth to accumulate enough resources to flower.) Finally, species that typically live longer than two years before flowering and dying are called *monocarpic perennials*. Some such species may live decades, undergoing an incredibly long period of resource accumulation while growing to great size.

Iteroparity can be seen as the norm, whereas semelparity typically needs to be explained by special circumstances. Starting with plants, *annuals* are quite common, and rather easy to explain. The annual life history is especially common in highly seasonal deserts with long annual dry periods that expose plants to extreme water stress. To be perennial in such environments requires the sorts of extreme adaptations we have seen in big, slow-growing, water-storing cacti. But an alternative strategy is to be a small, fast-maturing plant that compresses all of its active growth into the brief wet season, then produces *seeds* that spend the dry season in a state of *dormancy*. In this way, plants don't have to endure the full harshness of the environment while they are actively growing.

Weeds are a second class of plants in which annuals are prominent. These are plants adapted for growing in transient habitats that have been *disturbed*, either by natural phenomena like animal burrowing or by human activities such as agriculture. Soil disturbance allows a brief period for small annual plants to occupy an uncrowded bare site before they are pushed out by taller perennials that are better competitors. They can be considered *fugitive species* because they have to keep "discovering" new bare patches.

These two classes of annuals face very different challenges, but there are common features that favour the annual strategy. Both desert annuals and weedy annuals are adapted to take advantage of *brief windows of time* during which rapid growth is made possible by milder conditions. Both of them develop long-lasting *seed banks* in the soil, so that a large proportion of the population is underground at any one time. And both have elegant adaptations that selectively promote seed germination at the right time. Seeds of desert annuals are normally prevented from germinating by water-soluble germination inhibitors in the seed coats. These seeds germinate only after heavy rains have leached out the inhibitors, guaranteeing that the plants start growth when water is available. Seeds of weedy annuals, in contrast, tend to germinate only when they are struck by sunlight. This lets them get started right after there has been soil disturbance that brings some of them to the surface. Most of them remain in the seed bank, dormant and buried.

A third set of annuals includes many crop plants, particularly the three major grains: rice, maize, and wheat. Their wild progenitors were probably perennials. Here, the annual strategy has been developed through *artificial selection by humans* to maximize seed output, and to allow regular cultivation of the soil. Standard annual crop replacement in *tillage agriculture* requires fossil-fuel inputs and causes costly soil erosion, though, so some researchers are trying to develop perennial varieties of these crops. It remains to be seen whether those new crops can compete in the marketplace against the standard annuals.

Waiting a long, long time

It is harder to understand how long-lived semelparous perenniability can be an advantageous strategy. As described in lecture, plants such as agaves (“century plants”) grow for decades into a giant rosette of leaves, without flowering. Finally, some environmental trigger sets off flowering. All of their stored resources in the leaves and roots are re-mobilized into building a tree-like flowering stalk that can tower 10 metres high. Assuming that the flowers get pollinated, hundreds to thousands of fruits are produced, each packed with seeds. This desperate effort completely exhausts the plant, and it dies, earning the informal but apt name of “*big-bang* reproduction”. What circumstances make it advantageous to stretch out the pre-reproductive period of resource accumulation for so long? Various theoretical explanations have been offered, but there is no universal consensus. One attractive explanation is that the animal pollinators of such plants may have been disproportionately attracted to visit the plants with the largest inflorescences (i.e., larger flower displays). Such preferences, if strong enough, could confer reproductive advantages on plants that waited longer to flower, thereby storing up more

resources and being able to produce showier displays. This mechanism has some support.

The single, massive reproductive episodes of long-lived semelparous perennials lead to further adaptations regarding the synchronization of flowering among individuals in a population. Consider: waiting 70 years to flower would be a sad exercise in futility if the plant ends up flowering by itself, and there are no pollination partners to have sex with. Therefore, we might expect that plants might evolve to use the same cues to induce flowering. As I will illustrate in lecture, this certainly happens in the monument plant (*Frasera speciosa*) in Colorado. There, in most years, few if any plants bloom. In some years, however, almost all of the plants that have grown large enough will flower and die. Their highly conspicuous flowering stalks can reach 2 metres, so the subalpine meadows look very different in *Frasera* years. Long-term studies by David Inouye suggest that the cue for a big *Frasera* year might be *heavy rains four summers before*. Certain Asian bamboos are the epitome of synchronization; plants from one seed crop all flower in the same year, decades, later, even if they have been transplanted to another part of the world. As far as I know, the timekeeping mechanism for this adaptation remains unknown.

Synchrony, masting, and predator satiation

Extreme synchronization may have additional consequences that could also act as selective forces in maintaining it. Plants lose many seeds to predators, especially insects and rodents. If a plant population produces a constant crop of seeds every season, the populations of seed predators can build up to levels that are sufficient to kill nearly all the seeds as they are produced. A steady-state balance could be set up that would be detrimental to seed survival. But if seed production is synchronized across a plant population so there are occasional years of high seed production interspersed among more common years of no production, then the predator populations are kept low during the years of dearth. When a big seed crop comes along, the predator populations become *satiated* before all of the seeds are eaten up. Although the predator populations will build up during a year of abundance, they will crash the next year when seeds become scarce again. When Asian bamboos flower and set seed, rodent populations overshoot their carrying capacity so extravagantly that they migrate out of the seed-depleted bamboo forests and cause serious damage to human stores of rice. Before that happens, though, many seeds have been spared long enough to germinate and become the next generation of bamboo seedlings.

Population-wide synchronization of seed production is not restricted to semelparous plants. In the deciduous forests of North America, Europe, and China, numerous nut-bearing trees—especially oaks—produce heavy crops (called “mast crops” in these trees) roughly every 4 years. These species are regular iteroparous perennials, though. Studies in California by Walter Koenig suggest that predator satiation is probably a factor in masting by oaks.

“Big-bang” reproduction in *animals* is not very common, but you probably know one familiar and economically important big-banger: the salmon. Here, the fish begin life in freshwater lakes or rivers, migrate downstream to the sea, grow to adult size in marine waters, and then return to their natal river to spawn. Like an agave, a spawning salmon typically transfers so much of its tissues into gametes that it dies as a result. Another big-bang animal that is neither familiar nor economically important is *Antechinus*, an Australian genus of shrew-like marsupials. In this bizarre case, females are iteroparous while males are effectively semelparous. At breeding season, when food is scarce, the males go into such a frenzy of courting, fending off rival males, and copulating that they deplete their tissues and die of exhaustion. As might be expected, the females don’t get quite as worked up about things, and have the potential to live 2 to 3 years.

Dubious but effective mnemonic device for semelparity. I remember some professor claiming that the term semelparity alluded to the mythological character Semele. I can’t verify this, and it might have been a complete fabrication. But it’s so usefully dramatic that I’ll pass it on, anyway. Semele was a mortal woman who was enjoying an affair with the god Jove, who appeared to her in human guise. But Jove’s jealous wife Juno told Semele that she should ask to Jove to make love to her in his godly form, to get the full erotic impact. Jove reluctantly agreed to Semele’s pleading, but the divine encounter caused her to explode in flames. She was pregnant with Bacchus, whom Jove thoughtfully plucked from her womb as she perished. You can’t ask for a better image of big-bang reproduction! Bacchus, of course, grew up to be the god of wine, revelry, and undergraduate binge drinking, so that was a lucky save. The next time you are drinking, raise a glass to poor Semele. And another one to those male *Antechinus*. Perhaps you had better stop at that point.

Broadly applicable life-history strategies: the r-K dichotomy

Life-history characteristics do not vary independently across species. Instead, certain traits tend to co-vary, resulting in clusters of traits that are associated with either “fast” or “slow” life histories. Fast life histories are exemplified by fugitive annual weeds. They have high fecundi-

ty, low survival, short generation times, small but numerous seeds, good dispersal ability, poor competitive ability, and other characteristics listed in lecture slides. In a venerable dichotomy, such plants (and their animal equivalents) are termed *r-selected species* or *r-strategists*. Species with the opposite attributes (slow growing, good competitors, etc.) are called *K-strategists*. These names have aroused some discomfort. Some researchers don't like the term *strategy* for a set of adaptations because they feel that it implies some kind of teleology or advance planning by the organism. Agreed, it's important not to impute motives or goals to evolutionary processes, but I think we can use the well-established term "evolutionary strategies" without implying any end-directness. A second objection is that making up labels from the logistic parameters *r* and *K* might be taken to imply that these categories translate to particular values of those variables. They don't—the idea is simply that "*K-selected*" species show adaptations that would serve them well in high-density, competitive situations, and "*r-strategists*" do well in low-density situations. The essential point is that the parameters of the life table—the fecundity and survivorship schedules—do relate in a coherent way to the kinds of challenges species need to meet in different environments. We will come back to these characteristics when we consider ecological succession.

Thinking about the fast-slow or *r-K* axis presents an opportunity to outline two larger issues regarding functional classifications of this sort. You will doubtless encounter analogous issues throughout your career in science—or in any career with intellectual content. First, the human mind evidently loves sharp dichotomies and discrete categories. We feel comfort in the power of categories to organize and systematize information by tucking related items into pigeonholes. Nevertheless, most of the variability in nature is continuous. Categories blur into one another, and intermediate states exist between the poles of dichotomies.

Second, most categories are usually artificial and arbitrary, even though they may reflect nature well. For example, the British ecologist J. P. Grime felt that we needed *three* categories, rather than two, to organize life-history strategies in plants. His "C-S-R" *trichotomy* included **Competitive** species, which mostly correspond to "slow" *K-strategists* that compete well in *crowded* conditions; **Stress tolerators**, tough plants that grow slowly in situations that remain *uncrowded* because the physical environment is very harsh; and **Ruderal** species, which are fugitive-type weedy species that exploit transient disturbances to grow quickly in un-stressful sites that are *temporarily uncrowded*. We can all agree that Grime's three-way classification potentially *explains more ecological variation*.

tion than the binary *r-K* model does. Most plant ecologists would consider it a *better* conceptual formulation. But wouldn't a four-way classification be better still? If so, why not five? One could keep subdividing categories ever more finely, until each species occupied its own unique category. At that point, categories no longer have any added power to order things according to their affinities. Categories become nothing more than extra labels for the species themselves. A category's usefulness depends on its ability to bring together a substantial set of similar entities. The point is that *categories are usually arbitrary constructs*, and the level of subdivision is a matter of taste or practical convenience. We will return to some of these general issues when we treat plant community ecology.

Stochasticity revisited, with reference to modeling extinction risk

So far, all of the population models we have considered have been *deterministic*. What does that mean? The parameters that determine births and deaths are treated as *constants*, which means that the outcome is completely *determined* by the starting conditions. For example, if you run a Leslie matrix projection model that always starts with 100 newborns, you will always get exactly the same population trajectory. Clearly, deterministic models can have heuristic value, but we know that nature doesn't work this way. Chance variation will inevitably intrude, and there are some good reasons for including realistic variability if we are trying to make realistic predictions. I introduced this before, but let's now consider it again in the context of a typical application in conservation biology.

Imagine, for example, that we wish to construct a life table for a population of an endangered species to decide how much trouble the species was in. We would have to make some estimates of n_x , b_x , and l_x . We could then build a life table using the mean values of our parameters, and we could calculate λ to see if the population would be expected to grow or not. Suppose that λ came out to equal 1.001. That would suggest that we had a growing population (sigh of relief!), but all of those parameter estimates would be subject to measurement error. It would be imprudent to rest easy about the fate of this population.

What would be better would be to estimate not only the average values of the life-table parameters, but also their variances. We could then run *many* projections of the population's growth trajectory, say a thousand, each time using the computer to draw parameter values randomly from a distribution with the specified mean and variance. Some of those trajectories would show growth, but in others, we would have drawn a number of low parameter values by chance, and the population would

decline, perhaps even to extinction. Now our output or our “answer” would comprise all 1000 runs, and we would ask questions about the statistics of that set of trajectories. In what fraction of the computer runs does the population go extinct? In how many cases do numbers drop below certain thresholds of interest? Such questions give us far better handles on the actual risk to the population than any deterministic answer could. Therefore, stochastic models have generally replaced deterministic models in real-world attempts to gauge the risk of extinction. You will encounter stochastic models again when I discuss metapopulation models of pikas in Part 7.

Part 7. Community ecology, primarily regarding animals

We now take another step upward in complexity to consider the collections of many species that comprise ecological communities. Community ecology is a multifarious, hydra-headed enterprise. In the context of BIO120, we can grapple with only a few selected topics. I have tried to choose some pieces that are interesting in themselves but also have some lessons to impart about the way science proceeds.

Scope of community ecology: definition and properties of interest

The most inclusive definition of an *ecological community* is simply *all of the organisms* or the *biotic* entities in some spatially defined locality. (If we also included all of the *abiotic* components, such as atmosphere, weather, hydrology, mineral nutrients, etc., we would be defining an *ecosystem*, but we will defer most ecosystem material to BIO220.) Hardly anyone tries to study a whole community, because hardly any biologists have the expertise to work with organisms ranging from bacteria and fungi to vertebrates. Therefore, most studies in “community study” consider only a *subset* of the species in a local ecosystem. Subsets might be *taxonomically-defined* or *functionally-defined*. For example, we could consider the bird community of a mountainside in Arizona, or the soil decomposer community of a southern Ontario maple forest. Frequently, the restriction goes further, down to the *guild level*. Organisms share a guild if they have similar *functional niches*. For example, we’ll mention a classic guild-level study by Robert MacArthur in which he studied a set of closely related warblers, all of which eat insects (mostly caterpillars) in spruce and fir trees.

What properties are addressed by community-level studies? Well, just as population ecology is concerned with abundances and age structure, community ecology is concerned with the number of species present (usually termed *species richness*) and the abundances of each of them. Usually, communities comprise a few species that are very common (sometimes called dominants), and many more species that are rare. The distribution of abundance and rarity determines the *species diversity* of the species assemblage. For example, a community with ten equally abundant bird species would be considered more diverse than one with one common bird species and nine rare ones. There is a great deal of theoretical and empirical work on the patterns and determinants of species diversity, but we will defer most of that to later courses in ecology.

Thought question on estimating species diversity: Suppose that you start identifying the species identities of trees in a forest. As you walk around, looking at more and more individual trees, you should find more and more species. How would you expect the cumulative number of species to be related to the cumulative number of individuals that you have identified? To be specific, if you graph # of species (y-axis) versus # of individual trees (x-axis), what shape would you expect the curve to have? How would you decide when you had an adequate estimate of the true total number of species in the forest?

Note: If you find this hard to figure out in the abstract, you could actually do the exercise, assuming that you have access to a forest and know how to identify trees. Most of you will find that impractical. If so, you should be able to find an analogous situation. For example, stand at the St. George subway entrance and ask passing U of T students what college they belong to. Graph the cumulative # of colleges as a function of the cumulative # of students polled. How many students will you have to poll to find all of the colleges? Would you get a different graph at the Museum station?

Other properties, anticipating Part 9

Ecologists occasionally delve into functional biological properties of communities like *primary productivity* (e.g., how much plant growth occurs per area per time) or biochemical variables (like nitrogen uptake or gas exchange), but those include enough abiotic components that they are usually considered *ecosystem properties*. We won't treat them in this course.

In addition to species richness and diversity, community ecologists are interested in other properties. These include the *stability* of species composition, *resilience* in the face of stress or disturbance, the types of *changes* that occur through time, and the kinds of *interactions* that link the species. Interactions include overtly hostile encounters, such as *predation* and *competition*, and apparently benevolent connections such as *mutualism*. Traditional ecologists are particularly interested in *trophic interactions* (that is, who eats whom—what is the structure of the food web?); evolutionary ecologists are especially interested in how species exert selective pressures on each other, and whether they might *coevolve*. This material will see us through to the end of the ecology segment of BIO120.

Two traditions of community ecology

Unsurprisingly, the traditional separation of botany departments and zoology departments within universities led to very different directions being established in the early days of community ecology. Neither of these traditions was fully satisfactory, in my opinion, and it is instructive

to compare them. Their shortcomings were quite different, almost in a complementary manner. Briefly, plant ecologists tended to start from *descriptive* data. Although plant communities are discouragingly *big things with slow dynamics*, it was easy for researchers to go out into the field, delineate some study plots, and identify what plants were there. Plants hold still while you count them, a delightful property in a study organism. Such studies dominated the field from about 1900 to 1960. This field was *data-rich!* The ecologists looked for patterns in the data, which led to some elaborate theoretical frameworks aimed at explaining the patterns. The ideas were clever and attractive. It is virtually impossible to do decades-long controlled experiments on whole forests, however, so the theories stubbornly resisted testing. As we will see in Part 8, the self-correcting nature of science finally did kick in—as it reliably does—but it took an embarrassingly long time.

Animal community ecology did have a parallel trend of descriptive studies, but the more dominant trend came from extending theoretical population ecology from single-species models (like the logistic) to *multiple-species models*. That orientation lent an imposing mathematical sheen to the field in general, but the emphasis on N meant that the field was always *data-poor*. Because it is so hard to actually track animal population sizes in the field, the theories were built up on the rather inadequate foundation of a modest number of studies carried out in small containers, with simplified “communities,” under controlled laboratory conditions. Those “microcosm” studies were elegantly performed, but they simply didn’t capture some of the important forces that determine community composition in the real world. We will look at the evolution of these models, and the lab studies that informed them.

Two-species version of the logistic equation: intraspecific and interspecific competition

When investigators wanted to consider how two-species populations might compete with each other, it was natural to turn to the logistic model. In the years between the World Wars, Alfred Lotka and Vito Volterra developed such models, both for competition between two species and for predator-prey interactions. Recall that the logistic equation modifies exponential growth by adding a braking term that slows down growth as more individuals of the species use up resources. That braking represents *intraspecific competition*. To turn the logistic into the *Lotka-Volterra competition model*, we just have to add a second braking term that accounts for *interspecific* competition exerted by a second species. To make this happen, we now need two simultaneous differential equations, one for Species 1 and one for Species 2. We also need to add subscripts to the r , N , and K values to show which of the two species we are talking about.

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{K_1 - N_1 - \alpha_{12} N_2}{K_1} \right)$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(\frac{K_2 - N_2 - \alpha_{21} N_1}{K_2} \right)$$

This pair of equations also includes a new parameter α , the *competition coefficient*, which has *two* subscripts because it refers to the effect that one species has on the other. It is essentially a scaling factor. Look at the numerator of the fractional part of the equation for Species 1: the term $-N_1$ shows how adding individuals of Species 1 cuts into the carrying capacity of Species 1. This is familiar from the logistic. Recall that the logistic equation effectively models an environment in which *resources for population growth* are supplied at a certain rate, and organisms use them up. The term $-\alpha_{12}N_2$ shows how adding individuals of Species 2 further cuts into the carrying capacity of Species 1. We need to add the competition coefficient because there is no reason to expect that the two species use resources in exactly the same way. If an individual of Species 2 uses only half as much of Species 1's carrying capacity as an individual of Species 1 does, then α_{12} would be 0.5. To keep track of the double subscripts on alpha values, memorize the phrase "on-by." That is, α_{12} is the competitive effect *on* Species 1 exerted *by* Species 2. α_{21} is the competitive effect *on* Species 2 exerted *by* Species 1. If α is large, interspecific competition is important; if it's small, there is not much effect.

What is the outcome of competition according to these equations? At this point, you might expect that I would offer an explicit solution of N_1 and N_2 as a function of t , as I did for the logistic. But I can't. (I have a memory of reading somewhere that someone proved that a solution *exists*, but that it hasn't been found.) That seems a bit deflating—where is the vaunted beauty and power of mathematics now?—but it's not really an obstacle to understanding how competitive outcomes depend on the parameters of the model. We may not be able to plot the precise trajectories, but we know how things will end up.

Here are the conditions for the four possible outcomes:

If $K_1 > K_2/\alpha_{21}$ and $K_1/\alpha_{12} > K_2$, Species 1 will always *outcompete* Species 2. "Outcompete" means that, regardless of the starting numbers, N_1 will go

to K_1 as N_2 goes to 0. Species 1 reaches equilibrium at its carrying capacity while Species 2 goes extinct.

If $K_1 < K_2/\alpha_{21}$ and $K_1/\alpha_{12} < K_2$, Species 2 will always outcompete Species 1.

If $K_1 < K_2/\alpha_{21}$ and $K_1/\alpha_{12} > K_2$, the two species will *stably coexist*, with both of them remaining below their K values at steady state.

If $K_1 > K_2/\alpha_{21}$ and $K_1/\alpha_{12} < K_2$, the competition is unstable, but the winner depends on the starting numbers. Whichever species gains an initial advantage will outcompete the other.

You don't have to remember these inequalities, but you do need to know the four possible outcomes. You also have to understand this: *the condition for stable coexistence is that each species inhibits its own population growth more than it inhibits the growth of the other species' population*. That should be intuitively comfortable. Under these conditions, each species' growth exhausts its own carrying capacity more than it exhausts its competitor's carrying capacity. Yet another way to say it is that intraspecific competition must be stronger than interspecific. Small α values generally favour coexistence, while large ones tend to produce extinctions of inferior competitors.

Extension to more than two species

As you might have anticipated, the L-V competition model can be generalized to consider multi-way competition by multiple species. All one has to do is to write a differential equation for each species, and to subtract more braking terms including more pairwise α values. For example, the equation for Species 1 in a three-species community would be:

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{K_1 - N_1 - \alpha_{12} N_2 - \alpha_{13} N_3}{K_1} \right)$$

If you try to describe a large community in this way, the subscripts get a little tedious, and the stability properties are much more complicated, but it's still possible to say some things about the mathematical properties of the resulting set of simultaneous equations, using the techniques of linear algebra. Indeed, the set of all pairwise α values, arranged as a square matrix, became identified as "the community matrix," with considerable power to predict the population outcomes for a whole set of interacting species. The mathematical properties are far beyond the scope of an introductory biology course, but the model does serve as a

springboard for asking some biological questions, too. The obstacle to actually applying this model to real communities, of course, lies in getting accurate estimates of the pairwise α values. The rigorous way of doing this would require a large number of experiments in which population sizes of some species were manipulated, and responses of other species measured. Such experiments are effectively impossible to conduct in real communities.

Well, what about real organisms?

Also in the 1920s, G.F. Gause was investigating competition with various species of the aquatic protozoan genus *Paramecium*, not in real ponds but in laboratory cultures. As I'll show in lecture slides, populations of these microscopic protists tended to behave in ways that resembled predictions of the L-V competition models. Certain pairs seemed incapable of coexisting, with one of them going to deterministic extinction. Others tended to equilibrate, as if they met the criterion for stable coexistence. Another influential set of "container" trials was conducted by Thomas Park, who worked with two congeneric species of flour beetles. These species are so-called "stored-product pests" that commonly infest stores of flour. Park showed that competitive exclusion was the rule; he went further, and varied the environmental conditions (e.g., temperature, moisture) under which he maintained the containers of flour. Under certain conditions, one species tended to win; in other conditions, the other beetle prevailed.

Park's experiments therefore clearly indicated the importance of *abiotic* conditions in determining the outcome of biotic interactions between species, but also tended to reinforce the idea that it was unlikely or difficult for two competing species to coexist. In fact, that idea became enshrined as the *principle of competitive exclusion*, which is usually phrased as "two species that compete for the same resources cannot coexist for long." The motivating idea is that one of the two species will have to be at least slightly better at growing on that shared resource, and that the more efficient species will be able to increase its population at the expense of the less efficient competitor.

Some ecologists considered the principle of competitive exclusion a "law," and this idea had a disproportionate influence on the development of ideas in community ecology. In large measure, this arose from arguments tying competition to the ecological niche.

(Note in passing: as of September 2014, the Wikipedia entry on competitive exclusion was very poor. It seems to have been written by someone more concerned with promoting a quirky personal viewpoint than with providing

an objective account. It has now (September 2017) improved, but still strikes me as sketchy. Wikipedia can be authoritative only if authorities are paying attention to it. Be careful.)

Association of competition coefficients with ecological niches

In his early formulation of the niche concept, Grinnell was already thinking that it was unlikely that two species would occupy the same niche. If they were *doing the same thing*, you would expect one of them to be better at it than the other. But there are plenty of situations in nature where organisms *do* seem to be filling the same functional niche, and still managing to coexist. Do such examples comprise true abrogations of the “law” of competitive exclusion, or are we missing something?

Asking that question led Robert MacArthur, a student of Hutchinson’s, to conduct one of the most influential studies ever. As previously mentioned, he asked *how so many species of insectivorous warblers were able to coexist* in their summer breeding habitats of northern coniferous forests. These habitats do not offer a large diversity of tree species or of food species, so these birds would not seem to have much scope for specializing into different niches. Nevertheless, good habitats—like the places in central New Brunswick, where I did postdoctoral research—support as many as 20 different species of these very similar birds! MacArthur looked for differences in how these species used their habitats. In his 1957 paper, he found numerous differences of various sorts, but one particular difference became legendary. With binoculars and stopwatch, he did “time-budget” behavioural studies of the sort used by “efficiency experts” to speed up industrial processes. He documented that the most common warbler species were usually *feeding in different zones* of the coniferous trees (see lecture slides for clarification). That observation offered at least a partial explanation of what the species were doing differently: they were engaging in finer-scale *niche partitioning*.

The impact of this study was tremendous, for a few reasons. First, it was quantitative. One could actually *measure* the similarity in the way that two bird species used their shared habitat. By applying mathematical indices to behavioural data, one could compute the degree of *niche overlap* on a scale of zero (no sharing of resources) to one (complete similarity). Furthermore, one could plausibly argue that those *measures of niche overlap should be equivalent to the competition coefficients* in the Lotka-Volterra model. Measuring alphas was, of course, a sort of Holy Grail. Finally, those estimates of alphas could be obtained by simply observing avian behaviour—which is something that birdwatchers loved to do already.

Consider just how big a breakthrough this was! This approach offered a sweet alternative to the impossibly painstaking approach of estimating alphas by measuring population-size responses to field experiments. Studies of niche partitioning became wildly popular, and theoretical models developed additional concepts such as the idea of *limiting similarity*—the maximum amount of niche overlap that would allow two species to coexist. Niche overlaps and limiting similarity were also thought to determine community *assembly rules*: If you envisioned potential colonists arriving on an island or a new patch of habitat, the only ones who would be able to colonize successfully would be those whose niches were sufficiently different from those that were already established. This heady viewpoint offered power. We could not only understand but also *predict* the species dynamics of communities. It had the added lustre of mathematical sophistication. In the late 1960s and early 1970s, it seemed that a whole generation of graduate students adopted the framework of resource partitioning for their dissertations in community ecology. I was one of them. Generally speaking, this was a Bad Idea.

Is a resource-partitioning viewpoint sufficient?

What view of community structure arises from this model? An *equilibrium* view. As in the simple logistic model, the L-V models conjure up an orderly sort of community, in which each potential member species has either equilibrated at a steady-state value or has gone extinct because it was outcompeted. Whether or not a species can coexist with a set of others should depend on *patterns of resource use*. Furthermore, those patterns should be inherent, fixed, species-specific properties that can be adequately represented by the alphas. This implies that populations grow until they bump up against the carrying capacities of the environment. Resources are chronically limited; therefore, potential members of the community will be sorted out by competitive exclusion so that only the ones with different niches remain.

There are some well-studied cases where this paradigm seems to hold pretty well. Probably the best example is the distribution of *Anolis* lizard species across the islands of the West Indies. These small insect eaters do seem to divide up habitats in patterns that reflect functional feeding niches. Some species forage on the ground, others use tree trunks, others use small branches in tree crowns, etc. Furthermore, these behavioural feeding specializations co-vary with sets of morphological characters that are related to the different mechanical challenges posed by the different substrates used by the different lizards. Which species co-occur on different islands does seem to depend (in large part) on their foraging niches. To cap it all, these animals are amenable to population-

level experimentation. On tiny islets, it is possible to remove species or to introduce them in particular combinations and see what happens. The overall message is that resource partitioning/niche differentiation does appear to be a potent factor in the distribution of anoles in the Caribbean.

In many other cases, however, the correspondence between niche overlaps and alphas is elusive. I can illustrate this from my own dissertation work from the mid-1970s. I was interested in the idea that species of flowering plants might compete for the services of pollinating animals. I chose to consider the plant communities of subalpine meadows in the Rocky Mountains. I measured two aspects of “pollination overlap” for all possible pairs of plant species: first, how much did the two species overlap in their times of bloom, and second, how much they overlap in the species of animals that visited their flowers. It was time-consuming to measure those things for a whole community, but not technically difficult. I am confident that I obtained very good measures of niche overlap.

If I had stopped there, my thesis would have been like many others of the period. I would have presented my “community matrix” of niche overlaps and done some hand-waving about how the observed differences among “pollination niches” must have been sufficient to explain why these plants could coexist. But I was influenced by some skeptical papers that questioned the conventional wisdom, and I therefore made *direct estimates of the pollination success* of many plant species by using fluorescent powders to quantify how much visitation they had received. If the prevailing theories were correct, the plant species that were experiencing the *most niche overlap* should have received the *worst pollination service*, but that was not the case. Over the whole community, there was *no correlation* between niche overlaps and pollination success. Some plant pairs *did* seem to be competing for visits, but other pairs seemed to be cooperating or *facilitating* each other’s visits. The animals were engaging in a lot of subtle foraging behaviours that affected how well the plants did—behaviours that I have spent the last 30 years trying to understand. But the take-home message for “community structure” was that my niche measures had no dependable or straightforward meaning. *Overlaps did not measure competition.* Numerous others were reaching similar conclusions in other systems, and the whole enterprise of measuring niches as a *shortcut to understanding competition* dropped out of vogue. It usually didn’t work very well.

What are some of the reasons why this once-popular approach ultimately fell from favour? There are many, but some of the most important are factors that tend to make real communities *non-equilibrium*. Populations are influenced by factors *other than competition*. Let's consider some of them.

Natural enemies: predators and pathogens

In most ecology textbooks, you will find sections on predator-prey interactions in addition to those on competitive interactions. Lotka-Volterra models figure prominently in most of those treatments. It is also commonplace to encounter analogous models of how infectious diseases affect host populations. Here, I choose to completely skip over the mathematical details of such models, and just cut to the chase regarding the primary take-home messages for community ecology. (BIO220 will introduce some analogous host-pathogen models, with particular reference to infectious human diseases in the context of evolutionary medicine.)

As you might imagine, a Lotka-Volterra model for the interaction of a prey population with a predator population involves two differential equations for the population sizes of the two species. There are also parameters that determine how encounters between predators and prey get translated into population change. These encounters are modeled similarly to the “mass-action” models that you have probably encountered in chemistry courses to describe reaction rates between chemical ions in solution. Ecological *encounters* increase with the population sizes of the organisms, just as chemical reactions increase with the concentrations of reactants. Each encounter boosts the predator population while decrementing the prey population.

The result, as will be graphed in lecture slides, is a repeated pattern of regular oscillations through time. When prey are abundant and predators are scarce, there are plenty of prey organisms for each predator to encounter; predators are easily finding plenty of food, and they reproduce freely. But the increase in predators drives the prey population down. Eventually, there are too few prey to support the large predator population; predators start to go hungry or starve, and the predator population drops. The result is an endless pattern of *lagged population cycles*. Graphs of numbers versus time resemble two sine waves, with the predator curve lagging behind the prey curve.

For the simplest L-V formulation, those oscillations continue forever, much like the stable limit cycles we saw in May’s models of single populations with overshoots. However, when some additional parameters

are added to make the models more realistic—for example, adding a carrying capacity for the prey population—the cycles tend to either *damp out*, resulting in both prey and predator leveling off asymptotically at stable values and coexisting indefinitely, or to become *amplified*, resulting in mutual extinction.

Do *real* predators and prey display long-running lagged cycles?

Not so much! Many investigators, including Gause, have grown predators and prey together in containers under controlled lab conditions. In these simple environments, populations do show a *tendency* for lagged cycles, but they don't persist for long. The usual outcome is that the predator species finds and consumes *all* of the prey in the container and then starves before more than a couple of cycles are completed.

These lab results can't be a suitable representation of what happens in nature, because long-term persistence of predators and prey is commonplace in the real world. In lab cultures, the coexistence of predators and prey can be extended by making the culture conditions more complicated. In particular, if physical structures are added that give hiding places or *refuges* for some of the prey, predators are no longer able to hunt down every last prey individual. Instead, when the prey population becomes scarce, the predator population declines *without having completely wiped out* the prey population. This is almost certainly a general case in natural systems. Natural habitats offer infinitely more nooks and crannies than a glass beaker on a lab bench.

A second difference is that natural habitats usually offer *multiple prey species* that predators can take. If so, when one prey species has been driven to low numbers, predators are likely to *switch* to another species that is easier to find. Like a physical refuge, this mechanism allows a prey population to recover after being driven to low numbers. If this mechanism is important, we would predict that predator-prey cycles would be most pronounced in two situations: (1) simple communities with low species richness, where predators have little choice, or (2) cases where predators are obligate specialists that can successfully consume only one species of prey.

There is some support for both predictions. For predators who are *potential generalists*, such as vertebrate carnivores, the most pronounced, long-lived cycles are seen in the low-diversity communities of the Arctic, as I'll discuss in lecture. We also see more cycling in herbivorous insects, which are frequently obliged to eat only one kind of plant because it is the only plant whose chemical defences they have evolved to overcome. (We will discuss this important phenomenon in Part 9.) In both

cases, the predator's fate is intimately tied to a single resource population—the opposite condition of having a well-balanced investment portfolio with fallback options.

Returning to community ecology, however, all of this discussion of predator-prey interactions resolves to one basic conclusion: it's reasonable to expect populations of many species to be *depressed below their carrying capacities* by the actions of predators (or parasites or diseases, which can produce similar dynamics). In that case, those populations will seldom be bumping up against K ; they will not be experiencing the shortages of resources that lead to competitive exclusion. In short, predation can in principle *prevent inferior competitors from being outcompeted* by superior competitors. It keeps competition for resources from reaching an equilibrium state.

Does predation actually counteract competitive exclusion in nature?

Yes, at least in some cases. In lecture, I'll describe a tremendously influential paper from 1966. Robert Paine described a long-lasting *field experiment* in which he continuously removed predatory starfish (genus *Pisaster*) from the “rocky intertidal” zone of ocean shores in the Pacific Northwest. The *intertidal* is the zone that is submerged at high tide but exposed at low tide. On sandy beaches, not much life can get a toehold there, but when the substrate is rocky, there are many species of plants and sessile animals that attach to the rocks. In this community of disparate organisms, *competition for space is keen*. Where Paine removed the starfish, the number of species present dropped over the years; the rocky surfaces became *dominated by a single species* of mussel. Other species became extremely rare or disappeared entirely. In a control area, mussels did not push out the competition. They remained just one component species within a diverse community.

What was happening? The mussels were superior competitors for space, capable of pushing out every other species and creating a mussel *monoculture* unless something restrained them from doing so. However, the starfish provided exactly that restraint. Mussels are tough, armoured animals with few vulnerabilities, but *Pisaster* is specialized for killing them. Pitting its specialized tube feet and unrelenting hydraulic force against the bivalve's rapidly fading adductor muscle, a starfish can grimly pull apart a mussel's shells and digest the helpless animal inside. By creating mussel-free patches of habitat, starfish continuously create empty spots for non-mussel colonists.

Other ways to prevent competitive exclusion

The rocky intertidal is a special case, and Paine's experiments were not perfect, but the general principle has been upheld in many subsequent experiments from numerous ecological communities. I'll mention some of these in lecture.

Predators, parasites, and diseases are especially effective mechanisms for driving populations below their carrying capacities, because all of these hostile forces do best when the species they attack become abundant. These *natural enemies* tend to act in a density-dependent manner. However, density-independent, abiotic factors can also contribute. For example, Paine and colleagues also showed that species diversity was lower in calm-water rocky intertidal areas than it was in areas that received more violent wave action. The mechanism? When storms blow in, the pounding surf slams floating logs or ice against the rocks. These missiles scrape off mussels and other species. The resulting patches of open rock surface serve the same function as the patches opened up by starfish predation. Of course, they can be much larger.

As a general rule, abiotic "*disturbance events*" can affect community composition by counteracting competitive exclusion. In Part 8, we'll look at analogous effects in the context of plant communities dominated by fire.

Thought question: Would you expect that intertidal species diversity would always increase monotonically with the violence of wave action, or would the relationship be more complicated? What factors would be involved?

Lessons from the rise and fall of niche-based studies of communities

In my opinion, the history of the niche-centred view of community composition is a valuable warning about the seductive properties of attractive theoretical ideas. The notion that easy-to-measure niche overlaps could serve as proxies for impossible-to-measure competition coefficients was a tempting shortcut. It was elegant and attractive, but its promise was simply too good to be true. There was an underlying logic, true, but that underlying logic could be valid only in certain circumstances—specifically, in equilibrial communities where interspecific competition for resources is paramount. These would be communities comprising K-selected species that are being maintained near their effective carrying capacities. When species populations are routinely knocked down by other factors (predators, disease, bad weather, physical disturbances), the importance of competition is diminished. The niche-centred view has other shortcomings—for example, alphas would

surely not be constants, but would vary with conditions (as Park showed). Also, we would expect alphas to evolve over time. Still, this viewpoint's biggest liability is its *assumption* that communities must be equilibrial. Please understand that this is not some special case. *All theories rest on assumptions.* One has to be skeptical about those assumptions. Ideally, they should be subjected to independent experimental tests. If that can't be done, one must try hard to imagine ways in which they could be wrong.

Unfortunately, seductive theories of this kind are common in all branches of science, and particularly in medicine. This is why I feel justified in spending so much time in BIO120 explaining why too many ecologists were insufficiently critical about this pretty confection of ideas. Scientists must be skeptical, never more so than when some proposal offers an easy way out. Interestingly, the current fad in community studies is to study the mathematical "network" properties of species assemblages. I personally am worried that this enterprise has no better foundation than "niche biology" did 30 years earlier. If any of you go on to careers in ecology, you can see whether my skepticism is justified. You saw it here first!

Beyond populations: moving to a more modern perspective on community properties

I will revisit the need for skepticism at the end of this text. To conclude Part 7 on a more positive tone, I will outline what I think is a really insightful advance in theory for modern community ecology. This approach tosses out the focus on population size that has motivated all of the models we have considered so far. We no longer count individuals at all! Instead, we consider the properties of a *metapopulation*, which is a collection of numerous smaller *subpopulations* spread across space. The concept is rooted in the idea that a region will contain numerous discrete *patches* of habitat. Remember my starting point for the course: everything is patchy. In a metapopulation model, a patch is a potentially suitable home for a subpopulation of the species under study. These suitable patches are spread around in a matrix of *unsuitable* habitat, in which the species can't establish or persist. Members of the study species have some ability, however, to *disperse* across unsuitable habitat to reach suitable patches.

For a metapopulation to have interesting properties, migration and colonization need to be rather infrequent. If patches are perpetually connected by steady flows of migrants, the metapopulation structure melts down into one big population. It is the isolation and disconnection that lets us treat subpopulations as quasi-independent entities.

Returning to the ideas of Part 2, patches represent areas that fall within the range of tolerance of the species. Matrix areas fall outside the range of tolerance, and are effectively death zones. Subpopulations are considered to be small enough that they can go extinct, leaving a patch empty. But empty patches can also be recolonized by individuals that are dispersing from occupied patches. Instead of paying attention to the numbers of individuals, we focus on the proportion of patches that are occupied. That proportion will depend on the dynamic balance between extinction rates and colonization rates. If we adopt the simplifying assumptions that all patches have identical properties (size, probabilities of extinction (e) and colonization (c)), and that colonization varies with the proportion of patches occupied, we can show that the equilibrium proportion of patches (\hat{p}) occupied is

$$\hat{p} = 1 - \frac{e}{c}$$

Obviously, species that have high colonization potential and low extinction rates will tend to be found in nearly all patches at equilibrium; those that are poor colonists and more subject to extinction would tend to be rare.

Thought question. Harking back to earlier concepts, would you expect *r*-selected species and *K*-selected species to differ in their equilibrium proportions of patch occupancy? Why or why not? What properties do these species display that might be relevant to metapopulation dynamics?

Incorporating more realism

In my view, the real utility of the metapopulation concept does not lie in the theoretical equilibrial properties of hypothetical systems that are defined—or should we say crippled?—by simplifying assumptions. Rather, the fruitful concept of *separate habitat patches connected by occasional dispersal events* can be built into simulation models that capture realistic aspects of natural systems. For example, real patches are likely to differ in size and suitability. Real patches occupy real positions in space; some are more isolated from sources of colonists than others. We don't have to pretend that those differences don't exist. If we can apprehend their potential importance, and measure them somehow, we can build them into models. In lecture, I will go through such an exercise with a stochastic model of pika populations that gives insight into the critical process called the *rescue effect*.

Rescue effects are tied up with differences in patch suitability that are often expressed as a “*source-sink*” dichotomy. The source-sink terminolo-

gy comes from physiology, where it is used to describe directional flows of substances from sites of production to sites of consumption. For example, in a photosynthesizing flowering plant, the leaves are a *source* of fixed carbon while a developing fruit is a *sink*. In metapopulation ecology, a source patch is a particularly large or suitable patch that maintains a growing subpopulation and serves as a *net exporter* of dispersing colonists. A sink patch is of lower quality, and is incapable of maintaining a subpopulation except through the immigration of colonists from other patches. It is a *net importer* of dispersers, and if dispersal were cut off, the subpopulation in a sink patch would dwindle to extinction.

However, a sink patch that is embedded in a healthy metapopulation, with source populations nearby, can remain occupied indefinitely. Migrants keep arriving to restock the otherwise doomed subpopulation. There are some obvious implications for land management and conservation. Suppose that you were tasked with preserving an endangered species in a region undergoing development that would destroy some habitat patches. If you had enough money to preserve only a few patches, it would be supremely important for you to choose source patches rather than sink patches. In addition to the pika model mentioned above, we will look at interactive simulations in lecture that should help clarify source-sink dynamics.

Thought question: OK, it is supremely important, but how would you go about determining whether a given patch is a source or a sink? Is it necessary to measure migration rates? That would be hard, given that migration is probably infrequent. Are there some proxy variables that might contribute useful information?

The previous arguments and examples make the point that a metapopulation model can capture realistic aspects of natural systems. As I have stressed since the first lecture, nature is inherently patchy. If we adopt models that assume nature to be homogeneous—for example, all of classical population growth—we must expect some Procrustean friction and lack of alignment. The metapopulation world view matches reality better, and therefore fits more comfortably. Rescue effects are likely to be hugely important to the maintenance of natural diversity, especially in habitats of marginal quality and at the edges of ranges.

Another mythological aside: In case you haven't encountered the story of Procrustes, he ran a roadside inn with one iron bed. He had the kind of uniformity-loving mind that would suit him well for "middle management" in the business world, but he was also a thug. If a guest were too tall for the

bed's length, he would shorten the guest by cutting off his feet. If a guest were too short, Procrustes would stretch him out on the rack until he fit properly. Procrustes provides a favourite metaphor for forcing entities into conceptual frameworks that distort their essential natures.

Emergent properties of metapopulations

In addition to providing a better fit to natural communities, metapopulation representations show that these *subdivided systems can display properties that homogeneous systems cannot*. This is a fundamental insight with application to many branches of science. The term “*emergent properties*” refers to attributes of a whole system that are not inherent in the individual parts of the system. Rather, they “*emerge*” from the structure that links the parts.

In lecture, I will go through two examples, one involving predator-prey models and one involving interspecific competition. To reinforce the take-home messages here, consider a predator-prey relationship of the sort I mentioned above: a simple lab culture of protozoa in a test tube with no refuges. In such situations, I explained that the predator population will typically hunt the prey population to extinction and then starve to death. Game over. With such dynamics, the predator-prey system is clearly unstable. The two species *cannot* coexist in a single test tube. But what if we have a rack of many test tubes, with occasional drops of water falling from above to allow occasional inter-tube migration through splashing? In such a situation, empty tubes that have undergone extinction of both predators and prey can be recolonized by prey. It's a bit like getting a new life in a video game. Eventually, some predators will reach those tubes and initiate double extinctions, but if dispersal events are just frequent enough, the game can continue essentially indefinitely. The two species *can* coexist in a metapopulation of many test tubes. That fundamental change in outcome is a property that emerges from the metapopulation structure. We would say that the co-existence is *global but not local*.

In an analogous way, spatial structure can stabilize interspecific *competitive* relationships that would be unstable in a homogeneous habitat. Sticking with our model of aquatic protozoa, suppose that two species interact according to the L-V models such that Species 1 always outcompetes Species 2 when they come together in a tube. Again, we see *local* instability. However, if Species 1 sometimes goes extinct after taking over a tube, and if Species 2 is better at dispersing to new tubes, we can once again envision extended global coexistence. In this case, the poorer competitor needs to be a better disperser. A fast-dispersing *r*-selected *fugitive species* can stay one step ahead of a pokier *K*-selected competitor.

And that game can go on and on. The *difference in life histories* allows prolonged coexistence at the metapopulation level. (If that proposition doesn't make sense, review the material on life histories.)

Thought question: *In the two examples described above, allowing dispersal from one tube to another is necessary for achieving global coexistence when local coexistence is impossible. Will an increase in dispersal always increase global coexistence?*

Part 8. Community ecology, mostly plants

A battleground of ideas

Plant community ecology developed very differently from animal ecology. Although the fields have been converging toward a more unified set of perspectives in the last few decades, the different intellectual trajectories are still evident and still instructive. In particular, early plant ecologists did not focus on dN/dt and on mathematical models, as animal ecologists did. Instead, they selected numerous study sites (or “stands” of vegetation) and made lists of what plants they found there. They then searched for informative patterns in the data, and picked arguments with their colleagues about what those patterns meant.

This was a field in which arguments persisted for long periods because plant communities are comparatively resistant to experimentation. You could gain some insight into animal communities by studying microcosms of protozoa in a test tube, but you couldn’t do the same with oak trees. The sizes and the lifespans of the organisms were prohibitive. In addition to those actual impediments to manipulative experiments, there was also the psychological impediment. Plant ecology lacked a *tradition* of experimentation. In those circumstances, pet theories proliferate. If an idea was consistent with most of the observational data, and if it seemed “attractive,” it was hard to dislodge it. Much the same thing happens in all fields where experiments are either undervalued or hard to perform, such as medicine and cosmology. In a field like physics, one crucial experiment can bring down elaborate theories. It is instructive to see what is required to topple entrenched ideas when we are restricted to observational data.

The nature of plant communities

Around 1900, the Nebraskan Frederic Clements was pioneering much of the descriptive study of plant communities. He began by studying the remaining stands of native prairie vegetation, and later went on to work in various other biomes. He gained great influence through books in which he propounded his interpretive theories, and he had numerous students who helped popularize his ideas. One of the things that piqued his interest was that certain plant species occurred together more often than one would expect by chance. This was not a new observation; it would have been a part of common folk wisdom. Nevertheless, Clements was in a good position to discuss the phenomenon rigorously because he sampled plant communities using methods that were particularly objective and quantitative for that time.

He was interested in patterns of *association*. For example, in eastern deciduous forests, if you find American beech trees in a forest, you will probably also find a lot of sugar maple and some other species like hem-

lock. On the other hand, if a stand contains many trees of black oak, you typically won't find much beech or sugar maple, but you will find white oak and hickories. These patterns of non-random co-occurrences are not absolute, but they are definitely real. With sufficiently objective sampling, they can be verified statistically beyond any doubt. Indeed, it is commonplace to speak of certain community types based on the tendency of the dominant species to associate with each other. In southern Ontario and the rest of the upper Midwest, ecologists have little hesitation in calling one stand a "beech-maple forest" and another one an "oak-hickory forest." The question is not whether associations exist; it is what they mean. Why do certain sets of plants co-occur or not?

Clements argued that recognizable and repeatable plant *associations* occurred because the constituent species were functionally dependent on each other's presence, much as the organ systems of a higher organism are interdependent. Such functional dependencies might arise, for example, if sugar maples cast a certain degree of shade that beeches needed, or if beech leaves decomposed to produce soils that were especially favourable to the growth of sugar maples—or detrimental to oaks. Whatever the specific causes were, they had to be rooted in the biological characteristics of the species. Indeed, Clements likened a plant community to a *superorganism*: not only were its parts (the component species) functionally integrated, but it passed through developmental stages in the same way that an embryonic organism matured into an adult. This was a sociological view that coincided with the emergence of human sociology as a fashionable academic discipline. Indeed, the term *phytosociology* came into wide use for plant community ecology. The trees, shrubs, and herbs of a forest must be playing particular roles, just as members of a human society have roles like doctor, student, or labourer.

More important than the superorganism analogy, however, was Clements's conviction that particular plant communities could be seen as belonging to *discrete community types*, just as organisms belong to *discrete species*. In his *typological* way of thinking, the "beech-maple forest" was a category, much like a Platonic ideal, and all real stands of such forests could be usefully thought of as belonging to that category or type. This led to higher taxonomies of vegetation types, just as species are grouped into higher genera, families, etc. In Europe, this idea took hold to such an extent that Latin binomial names were coined for vegetation types. For decades, many careers were devoted to erecting and refining these classifications.

The Gleason-Clements debate

Obviously, Clements had to offer an explanation for real-world patchiness. What if the south-facing side of a hill supported oaks and hickories, while the more mesic north-facing side was clothed in beeches and maples? He suggested that the true vegetation types would be separated by a special type of narrow transition zones he called *ecotones*.

Another American, Henry Gleason, begged to differ. In 1926, he published an alternative explanation for plant associations that did not depend on functional interdependencies. It is a very simple idea that should resonate with you if you remember my earlier emphasis on *ranges of tolerance*. Gleason suggested that closely associated species like beech and sugar maple tended to be found together because they just happened to have similar ranges of tolerance to abiotic factors. Ecological/sociological “roles” had nothing to do with it. Indeed, he suggested that plant species should be found wherever the physical conditions of weather and soil are suitable, except in cases where they have not been able to disperse there. Communities are open to any species that can grow in the local climate and soils.

Gleason cited his experience with floodplain forests along the Mississippi River. As one travels from the headwaters in Minnesota to the Gulf of Mexico, there is a complete turnover in tree species. None of the species found in Minnesota are found in Louisiana. But Gleason claimed that it was impossible to recognize any ecotonal areas of rapid turnover where *groups* of species dropped out and other sets came in. Rather, he suggested that each species reacted individualistically to the changing conditions along the Mississippi transect. Community composition changed as a smooth continuum, not as series of discrete types. (*Fifty Shades of Green?*) Gleason stated rather bluntly that a plant association is “*not an organism, scarcely even a vegetational unit, but merely a coincidence.*” In time, Gleason’s proposal became known as the *individualistic, continuum, or open-community hypothesis* of plant community composition, whereas Clements’s explanation is variously termed the *organismal, holistic, community-unit, or closed-community hypothesis*.

Well. Clements and his large group of associates attacked the radical new suggestion. They were vocal, and held the majority position. Gleason was made so unwelcome that he quit ecology and turned to plant taxonomy, where he became a leader. Sadly for the scientific method, however, the argument did not get settled as it should have been, by appeals to agreed-upon *evidence*. Invective and rhetoric held sway.

In the long run, though, Gleason's ideas were eventually vindicated, first by the independent research programs of Robert Whittaker and John Curtis in the 1950s, and later by a flood of other studies. Once again, the self-correcting nature of science caused better ideas to drive out worse ones. But why did it take thirty years?

I spend time on this "ancient history" because it offers timeless lessons that all young scientists need to understand. It's not like the case with niche overlap theory, where an inadequate model was adopted because it offered *simplicity*. In the organismal-individualistic debate, Gleason's conception was far simpler but was spurned nevertheless. What caused the differences? The Gleason-Clements case seems to have included a shameful element of bullying, but it also hinged on the importance of *objective sampling*. The members of the organismal school were basically good ecologists, but they really did believe that community types were discrete entities. That was an article of faith. Therefore, when they wanted to study the composition and functioning of a beech-maple forest, they went out looking for a *good example of a beech-maple forest* to serve as their study site—that is, they would *choose a stand that met their preconceptions* of what the typological beech-maple category entailed. By doing so, they unwittingly ensured that every additional study would reinforce their prejudices about the unique and distinct nature of vegetation types. What about the sorts of intermediate stands that Gleason pointed to as critical evidence? They would never even be chosen for examination because they weren't good specimens.

I hope you can understand why subjective choice of samples can assassinate understanding. Curtis wryly said, "The most important decision an ecologist makes is where he parks his car." (The masculine pronoun is indicative of the 1950s, not of sexism.) I won't elaborate on the data sets complied by Curtis and Whittaker, but both of them effectively looked at the composition of many stands along *ecological gradients*. That is to say, they collected quantitative data that were analogous to Gleason's anecdotal account of vegetation change along the Mississippi. I'll show schematic representations of Whittaker's *direct* gradient analysis in lecture. Curtis's *indirect* approach depends too heavily on multivariate statistics to make good lecture material, but it became the mainstream approach of descriptive community ecology. You can learn about it in EEB's community ecology course. Both approaches yielded the same fundamental message.

As Gleason claimed originally, plant species typically show *individualistic responses along environmental gradients*, and they react to abiotic conditions far more than to biotic conditions. Most variation in plant species composition is smoothly continuous rather than jerky. When we do see apparent ecotonal zones of abrupt vegetation change, there is usually an underlying abrupt change in abiotic soil factors. If not, the *apparent* abruptness is due to one or two large dominant species like trees. If we look at less visually dominating herbs or shrubs, we usually see individualistic distributions. Now, it is possible to find some Clementsian sorts of biotic interactions between plant species that do influence their joint distributions. Those effects are typically small in comparison to abiotic forces, however. Note that animal communities are likely to be more Clementsian in character. Specific biotic interactions like competition and predation, as discussed in Part 7, will often play stronger roles in animal populations.

Thought question: Because typical animals are mobile, all individuals of one species population can interact with all individuals of other species populations that move around in the same area. In contrast, individual plants tend to interact strongly only with their immediate neighbouring plants. How does this difference alter the nature of competition in these different sorts of organisms? What are some consequences?

Finally, the important part

After the introduction of gradient studies, a seemingly intractable debate about the fundamental nature of terrestrial plant communities could actually be settled satisfactorily. What it took was agreement on an *operational criterion*, specifically “Does community composition change continuously or discontinuously along an ecological gradient?” If both sides were willing to accept that criterion, then the problem could finally become susceptible to data rather than debate. Of course, those data needed to be gathered in ways that were relevant to the operational proposition. As long as ecologists focused their attention on detailed cataloging of a few, handpicked, exemplary sites, the big question would remain elusive. Only when people like Curtis and Whittaker zoomed their focus outward, enlarging the scale of their investigations to consider many objectively chosen sites, did the scope of the data match the scope of the question. Operational definitions are critical for *settling*—as opposed to fueling—scientific disagreements. By definition, they specify what operation you have to carry out to get an acceptable answer.

Optional mental exercise for students with some exposure to the history and philosophy of science: Whittaker was trained as an entomologist and Curtis as an orchid physiologist. Neither of them had been schooled in mainstream plant ecology. Do you think that their outsider status makes the Gleason-Clements controversy a good example of a Kuhnian paradigm shift?

Plant community dynamics: ecological succession

It is a commonplace observation that the species composition and the overall structure of plant communities change over time. Sometimes, this process of *succession* happens quickly enough to be studied by direct observation. For example, if you visit a particular vacant lot in Toronto each summer, you can easily see that the weedy plant community is different each year. Nobody is gardening these spots, so the changes represent spontaneous natural phenomena. My favourite way to think about it: “Mother Nature is a big girl. She knows how to dress herself.”

In other communities, such as forests, the turnover of species may be so slow as to be imperceptible over the active career of an ecologist. A healthy white oak tree can easily live 300 years; how fast can a forest be expected to turn over?

To study the process of change over time spans of decades to centuries, it is usually necessary to do some kind of “space-for-time” substitution. In what is usually considered the first study of succession, the pioneering University of Chicago ecologist Henry Cowles looked at sand dunes on the shore of Lake Michigan. He realized that wave action (driven by prevailing westerlies!) was continuously eroding sand from the western shore and depositing it at the south end of the lake, adding new shoreline over the centuries. This process created a series of parallel dunes that Cowles recognized as providing a time series or *chronosequence*. The oldest dune was the farthest inland, the youngest one nearest the water. This special characteristic of dunes gave Cowles a particularly elegant way to achieve a series of replicate habitats of various ages, but one can also assemble chronosequences by finding, say, a number of forests that have been developing on abandoned agricultural fields for different spans of time. Note that there are two kinds of succession: the more “pure” type is *primary succession*, where plants colonize newly created habitat that has not previously supported a plant community. Cowles’s dunes were undergoing primary succession. More common is *secondary succession*, where an existing plant community is disrupted or erased by a *disturbance*, such as farming, fire, windstorm, or logging. After the disturbance, the plants are disrupted but the soil is relatively intact.

You might think that successional change in community composition would be antithetical to Clements's conception of communities as discrete superorganisms, but he had that covered. He viewed the succession of a community as analogous to the *developmental stages* of an organism. To make that analogy cohere, he also posited that plant communities developed for a while, but then reached a stable "adult" stage that no longer underwent change. He termed that final stage the "climax" formation. Earlier, unstable stages were called "seral" stages. Indeed, Clements loved the idea of succession because the stages are largely predictable. That aspect is consistent with his analogy to embryonic development.

To make those abstractions more concrete, let's consider the example of secondary plant succession on abandoned farms in the deciduous forest biome of southern Ontario. Here is a simplified caricature. When an "old field" is removed from crop planting and cultivation, it is immediately dominated by annual weeds such as ragweed. Most of these weeds come from dormant seed banks already in the soil, and their rapid growth and flowering replenishes those seed banks. Within a couple of years, these are replaced by taller perennial weeds such as goldenrods. These literally overshadow the annuals and outcompete them for light. The perennial weeds hold the stage for a decade or two, but are gradually overtapped and shaded out by woody shrubs and tree saplings. The first tree species are usually short-lived, fast-growing, sun-loving "pioneer species" such as aspen, pine, and Manitoba maple; they are replaced in turn by slower-growing and more shade-tolerant species such as oaks, hickories, and red maple; and those species are eventually supplanted by extremely shade-tolerant species such as sugar maple and beech. (Although unlogged climax-type forests (now usually called "old-growth") are rather rare in our area, you can see all of these stages on a trip to U of T's Koffler Scientific Reserve north of the city.) You should recognize this sequence as essentially a progression from *r*-strategists to *K*-strategists.

Historical mnemonic digression: Recall that seeds of weedy annuals frequently germinate when exposed to light by soil disturbance. This was unforgettably exemplified during WWI by the swaths of red poppies that emerged from seed banks after agricultural soils were torn up by trenches, shell craters, and cemeteries. If you buy a poppy to remember the fallen soldiers, also give a thought to that flower as the start of secondary succession, a "healing process" of sorts.

In forest succession, one can distinguish a transitional seral stage from the stable climax stage by comparing the species composition of the mature canopy trees to the species composition of the young saplings in the understory. Under a canopy of oak, we are likely to find saplings of beech and sugar maple, so we can infer that the forest is still undergoing turnover. Under a canopy of beech and sugar maple, however, we find saplings of beech and sugar maple. That's the signature of a stable "climax" forest: *the tree species that remain are capable of reproducing under their own shade.* In Clement's characteristically dogmatic view, the climax configuration is the goal that *all forests* in a particular climatic region will eventually reach. That is, he would predict that all sites in southern Ontario (and in all of northeastern North America south of the boreal forest, for that matter) will eventually converge to beech-maple climax forests.

Clements reached this rather extreme conclusion partly because his superorganism concept worked best if there was a single "adult" mature forest type. That let him dismiss the obvious variation among different real forest stands as unimportant differences in developmental stage. Those disparate stands could still belong to the same *essential type* because they were all fated to converge to the same composition in time. Furthermore, this view made sense to him because he was convinced that the principal forces driving successional turnover were what he termed *autogenic* processes—processes that have more recently been termed *facilitative*. Facilitation occurs when one prevalent species *changes the environment* in ways that facilitate that species' replacement by another. For example, black oak seedlings are shade intolerant. When a bunch of black oaks grow to be full-sized trees and form a closed canopy, new black oaks can no longer thrive in the now-shady understory. Only shade-tolerant species can now succeed, so the species composition shifts when the existing canopy trees are replaced by the next generation. The "autogenic" tag emphasizes that the environment is changed *by the species themselves.* (Environmental changes that are driven not by the biotic community but by external forces—say, soil erosion or flooding—are called *allogenic*.)

The principal autogenic factors in plant succession are (1) those involved with soil development (which is most important in primary succession) and (2) competition for light (which is most important in secondary succession, because secondary succession occurs on soils that are already developed). We've already discussed light shading effects. For an example of soil development in primary succession, consider the newly formed sand dunes studied by Cowles. These are terrible places for

most plants because the substrate (pure quartz sand) doesn't hold water well, lacks nutrients, and is also unstable. It drifts with the wind. That means that slow-growing seedlings are susceptible to having their shoots covered up and smothered, or having their root systems exposed and dried up. The only species that does really well in those treacherous conditions is marram grass, a species that forms extensive root systems.

Marram has the beautifully apt genus name *Ammophila*, which means "sand-loving." It grows fast, and it grows even faster if it starts getting covered by sand. Once marram grass establishes a dense stand, it stabilizes the sand so it doesn't blow around as much. That change doesn't help marram very much, but it *does* allow other, slower-growing species to establish without being buried. Those species gradually outcompete the marram grass, whose speedy growth no longer provides a special advantage. In fact, marram suffers from the general tradeoff between growth rate and competitive ability. All of the species contribute organic matter to the soil through dead roots and shed leaves, and the once-sterile sand gradually becomes a complex soil with a meaningful component of humus to retain water and nutrients. Marram has paved the way for its own replacement.

That sort of *facilitative* force is exactly how Clements thought *all* succession worked. It is consistent with his analogy to developmental biology, but also with his shaky grasp of adaptive evolution. If the community were an organism, it makes sense that the role of one species might be to make things nicer for a succeeding species. But a community is *not an organism* (thank you, Dr. Gleason!), and natural selection doesn't act to produce altruistic chivalry towards other species. Yes, some genuine instances of facilitation do occur, but succession is also characterized by plenty of *inhibitive* effects, in which earlier species change the environment in ways that *retard* the success of later species rather than hastening it. They don't pave the way; they tear up the pavement and set up roadblocks. Most importantly, both inhibitive and facilitative effects are coincidental consequences of particular properties of the species. Natural selection has favoured those properties because of their *positive effects on individual fitness*, not because they produce certain kinds of successional change. To be more concrete, marram grass has a fast growth rate because faster-growing individuals of marram have produced more offspring than slower-growing individuals, not because fast growth contributes to autogenic turnover in succession.

Disturbance: toward a more inclusive view

In early plant ecology, the stately progress of succession toward climax was seen as the big story; fires, wind storms, and other disturbances

were seen more as pesky setbacks to the majestic progression. Now, we are far more focused on the particular *disturbance regime* that characterizes a habitat. In many ways, the pesky setbacks have now become the big story! We also realize that the idealized notion of climax is simply not applicable in many situations. For example, in the Mediterranean/chaparral biome, in many grasslands, and in drier pinelands, wild-fires occur so often that succession virtually never gets past its earlier stages. (I'll demonstrate those effects with computer simulations in lecture.)

Ecologists who study such ecosystems have moved from considering fire as an occasional, unnatural anomaly to recognizing fire as just another component of the environment. The plants and animals that live in such habitats show clear adaptations to fire. There are suggestions that some of the plants like pines and eucalypts may actually be under selection to produce flammable foliage, because the individuals of these species fare better as fire intensity rises. I'll conduct a lecture demonstration of an adaptation to fire that characterizes some pine species.

The classical climax idea also fails to apply to the huge boreal forest biome. There, succession from various starting points tends to lead fairly quickly to closed-canopy forests of spruces and fir, but these forests are so dark that there is no understory of young saplings waiting to replace the current generation of canopy trees. This forest is the endpoint of succession—it is as far as things go!—but it is not a stable, self-perpetuating climax of the sort we envision beech-maple forests to be. Instead, the canopy trees just get older and taller until they are vulnerable to being toppled by windstorms. Their root systems are vulnerable to fungal attack, and their shoots are rendered more susceptible to herbivorous insects. For these coniferous trees, an important defense against insects lies in the pungent volatile terpene compounds in their sap. When an insect probes a young tree, it gets a pressurized squirt of this nasty stuff in its face. Sap pressure falls as trees age, though, so insects seek out older trees as better dining spots. In drier areas, fire may also become a factor.

All of these factors mean that an old boreal forest tends to just run out of steam and fall over at some point. When that happens, a rather conventional secondary successional sequence starts with shade-intolerant pioneers in a rather open environment. A stable, self-perpetuating endpoint? Not even close.

Gap-phase succession

Another recent advance in our thinking about succession is an appreciation of the importance of *spatial scale*. In New England and eastern Canada, the most consequential ecological phenomenon in the 1600s through the early 1800s was the arduous conversion of native forest to farms by European settlers. In the late 1800s and 1900s, it was the abandonment of those farms, as farmers realized that the deep black soils of the midwestern grasslands were far better for growing crops. Universities were also concentrated in New England, and the science of ecology emerged and took shape during that farm-abandonment phase. As a result, many of the early studies were on the spatial scale of a standard 40-acre farm field. Their temporal scale tended to cover short, early stretches of secondary succession ranging from 0-100 years. Ecologists got accustomed to thinking of succession as a stand-level process.

For example, when I arrived at Stony Brook University in 1980, I set my ecology lab class to studying the oak forests from which the campus had recently been carved out. These were well-developed forests with a mostly closed canopy about 20-25 m high, but the oldest trees were only about 80 years old. We occasionally encountered old pieces of fence wire and a few rotting fenceposts. The forest floor had the telltale smoothness of a once-plowed surface, with only a few pits and mounds caused by the root balls of trees uprooted by past hurricanes. The recent agricultural heritage was obvious. It was easy to envision the operating farm—several candidates for the most likely farmhouse were still standing on nearby roadsides—and easy to imagine the date at which the farmer packed it in. At that point, a large swath of land entered secondary succession, all at once.

What a different feeling one gets when entering an old-growth tropical forest without a history of agriculture! It feels as if the tall, closed, dark forest has been there forever, a perfect exemplar of an unchanging climax. But! Every so often, you encounter a place where one of the giant canopy trees has blown down. Because tropical trees are frequently tied together by networks of strong woody vines (lianas), one falling giant will usually pull down a number of others. The resulting gap is a bright sunlit patch of green chaos, often as big as a city block. There is a quick scramble to see which plants can take advantage of the precious sunlight. Seedlings and sucker shoots spring up from the tangle of dying branches on the ground, while intact trees on the margin of the patch extend branches laterally into the sunny opening. If the patch is small, inward growth from the surrounding trees may close it off before any seedlings have a chance. Big patches provide more complicated dynam-

ics, including opportunities for shade-intolerant pioneer species to recruit from seed.

Tropical ecologists popularized the study of “*gap-phase*” succession because this phenomenon grabs one’s attention so forcefully. On a closer look, that dark, closed forest that seemed to be a perfect climax formation probably has a substantial proportion of shade-intolerant pioneer species in the canopy. Those trees are the winners of fairly recent races to fill in gaps. The forest is *not* uniformly old-growth. Rather, it is a complex spatial mosaic of many patches of different ages.

Actually, virtually all vegetation will have that mosaic character, with patches of various sizes in various stages of recovery from disturbances. That character is harder to see in old-field succession because the fine-scale pattern of single-plant replacements is overshadowed by the strong pattern of abandoned farms at the scale of 40-acres and up.

In summary, we have come around to conclude that the *nature of succession* is not determined so much by the “developmental biology” of quasi-organismal stands toward a climax configuration as it is by the individualistic responses of a species pool to a *disturbance regime* characterized by certain types, frequencies, and sizes of disturbances. Those disturbances not only provide *ecological opportunities for colonizing species*, but also exert natural selection that builds *adaptations for surviving or recovering* from disturbances. Some of those adaptations, such as heightened flammability, loop back to affect the probability or intensity of the disturbance regime. Because of this, it won’t be very useful to consider *succession* as some unitary, fundamental *process* with a set of universal theoretical *properties* that we can catalog, study, and theorize about. Properties that emerge in one region or biome may not emerge elsewhere.

For example, until at least the 1970s, influential textbooks claimed that plant species diversity should increase as succession proceeds toward climax. The idea was that the vegetation should become more complex as it develops, creating more niches for more species. This might even be true in some tropical rainforests. Empirically, however, many other systems violate this “rule.” In southern Ontario, for example, succession from an old field to an old-growth forest causes an *initial increase and then a later decrease* in species diversity. Whether a dark, late-successional forest might “offer more niches” is dubious, but it’s also immaterial: the regional pool of species offers relatively few shade-

tolerant species that can thrive in those conditions. There is simply a longer list of plant species that can survive in sun-dappled oak forests.

In lecture, we'll explore simple models of succession as a stochastic process driven by fixed probabilities of species replacements. In these models, too, species diversity is low in the pioneering stages following a disturbance. In the absence of disturbance, diversity rises and then falls, as succession proceeds to a low-diversity climax. To maintain the highest level of diversity requires an intermediate intensity of disturbance. We have seen analogous results before, such as the effects of wave-tossed logs on the diversity of rocky intertidal communities.

So, plant communities change, but their ways of changing tend to be as individualistic as their species compositions. To condense the above arguments to a “sound bite,” succession isn’t an entity with a set of properties. It is a diverse *collection of outcomes* that emerge from interacting properties of climates, soils, and organisms. If there is any broadly applicable theme, it is the *intermediate disturbance hypothesis* that diversity is enhanced when disturbance is neither too common nor too rare.

Thought question: As we discussed in the life history section, seed dormancy in desert annuals is often released by exposure to water. In agricultural weeds of mesic forest biomes, the releaser is sunlight. In chaparral plant species, seed dormancy is frequently broken by exposure to smoke. Do these properties of seeds serve similar adaptive functions in different biomes? What are the similarities and differences?

Part 9. Trophic relationships and plant-animal interactions

Having touched on the earlier phases of intellectual development in community ecology, I wish to conclude with some topics that are more current. The concept of trophic structure is a way of considering community organization that is based not on the population ecology of the constituent species, but rather on the ways that they acquire energy and nutrients: what do organisms eat? And who eats them in their turn? All organisms occupy some “*trophic position*” within what is often termed the “*food web*” of a community, based on their source of high-energy carbon bonds. Photosynthesizing green plants make their own from CO₂, and are therefore termed *autotrophs* or *primary producers*. Collectively, they constitute the *first trophic level* of a community or ecosystem. All of the other species are called *heterotrophs* or *consumers*. The *second trophic level* comprises herbivorous animals that eat plants; these are *secondary producers* or *primary consumers*. Carnivores that prey on live herbivores would be *tertiary producers* or secondary consumers occupying *trophic level three*, and so forth. Animals that consume from more than one trophic level are *omnivores*; consumer organisms that extract their carbon from dead organisms are called *detritivores* or *decomposers*.

Again, G. E. Hutchinson (along with his students and others) pioneered the study of the trophic structure of natural communities. The breathtaking intellectual novelty of this view lay partly in taking the point of view of a joule of energy, or an atom of an element. Hutchinson asked how this *indestructible little bit of something* moved from the abiotic world into the biotic world, how it flowed from organism to organism, and how it left the biosphere to pass back into the non-living world. In the case of *energy*, the beginning and ending of the story are predictable and ultimately boring: energy comes to the earth from electromagnetic radiation beamed out into space from the sun, and it leaves in the same form, being beamed out into space from the earth. In between its arrival and departure, however, it is the power that drives miracles. That includes you.

Turning to *matter*, the earth gets a small amount of input from meteorites and cosmic dust, but it is mostly a *closed system* in which energy propels atoms through characteristic *biogeochemical cycles* that depend on their elemental chemistry. Some elements don’t have much commerce with the biosphere. Of the others, some are highly mobile, highly reactive, and closely tied up with the metabolisms of organisms. Oxygen, nitrogen, and carbon are examples. Others, such as phosphorus, spend

most of their time in stable, insoluble, boring abiotic forms like sedimentary rocks at the bottom of the ocean. Occasionally, though, they emerge to be *recycled* through organisms for some exciting biotic adventures before returning to their slumbers like chemical Sleeping Beauties or Brünnhildes. During those lively interludes, the two dozen life-making elements combine to form all of the individuals of all of the 8.7 million species the planet supports. That includes you.

To say that there is a lot going on here would be a comical understatement, but most of it falls under the category of ecosystem ecology. Interestingly, the study of ecosystem ecology received a huge impetus in the 1950s from the development of nuclear weapons. When everyone expected nuclear war to break out and spray carcinogenic isotopes into the environment, it seemed very important to know where those radio-nuclides would end up. The study of elemental flow rates and pool sizes became central to ecosystem ecology. We'll defer highly physical processes like biogeochemical cycles and energy transfers to later courses in EEB or ENV. Here, we will stick to some selected aspects of trophic ecology that are closely linked to whole-organism biology.

Why is the world green?

A somewhat trite but utterly classic starting point for considering trophic ecology is a 4-page argument published by Hairston, Smith, and Slobodkin in 1960. It's so famous that it is universally referred to as HSS. (Slobodkin was a Hutchinson student, also a colleague of mine at Stony Brook.) They argued thus: in most terrestrial communities, there is a lot of plant material (primary productivity) that is not consumed by herbivores. Therefore, the second trophic level (herbivores) must not be limited by their food supply, and must instead be limited by their consumers (predators, the third trophic level). Despite its beguilingly simple logic, the HSS proposition for "why the world is green" generated controversy. The task of evaluating its soundness has been assigned as a brain-teaser to generations of students.

One criticism of HSS lies in an alternative explanation for the evident "greenness" of the world: perhaps not all of those unconsumed leaves are edible. We'll turn to that shortly. Furthermore, to the extent that the logic holds, the arguments only apply to *entire trophic levels*. Perhaps the second trophic level is in fact limited by predation from the third trophic level; that doesn't mean that all of the *species populations* in the second level are limited in that way. The implication is that the trophic-level proposition could be true but in a biologically uninformative way.

One branch of research to stem from these ideas was graph-theory research into the topology and “network properties” of foodwebs. As mentioned earlier, I’m skeptical of such approaches. I think that the data on “who eats whom” are generally inadequate. Observations are too scarce for rare species. Furthermore, in very recent years, it has become possible to quantify on what trophic level an organism is operating by measuring the ratios of stable isotopes in its tissues. Those numbers don’t always concur with observational data, and putative carnivores almost always turn out to be omnivores that fall between numerical trophic levels. I won’t say more about this subdiscipline, except that it is evolving rapidly with the advent of stable-isotope studies.

A second branch from the stem of HSS is the study of what have become known as “trophic cascades.” These are measurable, important, *indirect* relationships between two trophic levels that are mediated through a third level. Although HSS did not use the term trophic cascade, the defining example is their proposition that an increase in level-3 predators should cause an *indirect increase* in level-1 plants through the mechanism of the predators *directly depressing* the intervening level-2 herbivores.

Trophic cascades are interesting and important not only from a theoretical viewpoint, but also with regard to practical management of real-world populations. If communities are actually integrated in ways that cause a change in one species population to register in other species two or more trophic levels away, that is important to know. And unlike most documentations of “foodweb structure,” hypotheses about trophic cascades are frequently susceptible to experimental tests. The tests typically involve removing animals operating at the third or fourth trophic levels and then looking for responses at lower levels. Significant responses to the removal experiments would establish “top-down” trophic cascades. One can also look for “bottom-up” effects by removing primary producers.

Admittedly, such tests are practical only in rather simple communities, and are typically limited in spatial extent. In lecture, I’ll describe how ecologists have sought out special circumstances to make experiments possible (and sometimes replicable). These include working in little freshwater ponds, or on tiny Caribbean islets where it is actually possible to find and remove all of the lizard predators. The restricted spatial scale of these experiments also tends to ensure a low species diversity. On the plus side, these experiments demonstrate that trophic effects *can* cascade through systems in interesting ways that would be hard to imagine in advance. For example, before Tiffany Knight’s experiments,

few people would have thought the presence or absence of fish in a pond would have any effect on the pollination of flowers of plants growing nearby. That sort of demonstration usefully reminds us that natural populations are connected in subtle and improbable ways. On the minus side, restricting experiments to a single isolated patch banishes the possibility for metapopulation structure and interpatch migration. That in turn eliminates the possibility that responses could be tempered by rescue effects. This particular sacrifice of reality for simplicity could be very costly.

To produce truly replicated experiments, several investigators have set up aquatic “microcosms” of protozoa (much as Gause did) or larger “mesocosms” in big containers like tanks for watering cattle. There is currently great enthusiasm for mesocosm experiments on the scale of a cubic metre or so. The hope is that they might let us combine the power of replicated experiments with enough realism to produce results that adequately mimic natural systems.

Thought-experiment questions: How could you determine whether mesocosm experiments adequately reproduce natural effects? Do you think that experimental ecologists have a responsibility to conduct such tests before they conduct mesocosm experiments?

One generalization is that *trophic cascades are more likely to be found in aquatic systems* than terrestrial ones. This is probably due to the comparative absence of spatial structure in aquatic systems, especially small ones. Because the aqueous habitat lacks barriers to dispersal, the species are effectively “stirred up” in a homogeneous way that lets them interact more completely—more like chemical reactants obeying mass-action laws in a stirred beaker. They will behave more as populations and less as metapopulations. Terrestrial systems will tend to be more heterogeneous, and therefore more unruly.

The difficulties of herbivory as a factor in the generation of biodiversity

This section delves into my own field of interest, the evolutionary ecology of plant-animal interactions. When I entered the field as a graduate student, such studies were almost non-existent, but they have now become mainstream. U of T and a few other universities actually offer undergrad courses in such interactions (see EEB440: Ecology and Evolution of Plant-Animal Interactions). In BIO120, I don't have time to talk about my favourite topic, plant-pollinator interactions, but I must touch on some aspects of *plant-herbivore relationships* because these are critical to our main theme of biodiversity.

The lifestyle of eating plants is tough. Of course, *all* lifestyles are tough—the premise of this text is that the struggle for existence is universal! — but eating a diet of plants is tough in the special sense that it requires elaborate adaptations that flesh-eaters don't need. This is because plants have evolved to *defend* the tissues that are important to their fitness. Because plants can't run from herbivores or hide their photosynthetic tissue underground, their main recourse is to make it unpalatable. Note that I am considering *folivory*, or leaf-eating, here; some plant tissues, such as sweet fruits or floral nectar, are not defended because their adaptive function is *to be eaten*. They exist to reward seed dispersers or pollinators in relationships that are mutualistic, not antagonistic. They present a whole different story.

Plant defences are highly diverse; the broadest dichotomy is between *structural* and *chemical* defences. Structural defences include the thorns of cacti, the spiky hairiness of many leaves, the abrasive silica of grasses, or the physical toughness of sclerophyllous leaves. Chemical defences, in contrast, act as metabolic poisons or deterrents.

Plant species produce a vast array of chemicals that have no role in the basic processes of photosynthesis, respiration, and maintenance. Because those processes were considered the primary metabolic activities of plants, all of the other biochemicals tend to be called *secondary compounds* (or *secondary metabolites*; I fear that this term seems to be gaining popularity because it sounds more pretentious). In the early, descriptive days of plant biochemistry, there was considerable doubt about whether secondary compounds had any functions at all. One popular explanation among bench scientists was that they were waste products like urea. However, field naturalists objected that this didn't explain the spectacular chemical diversity of the compounds, or the fact that many of them had potent effects on the metabolisms of animals that ingested them. Many of these compounds had long been known and used by humans as medicines, as culinary spices, or as recreational drugs. In a lecture

slide, I list a number of *alkaloids* that have been used in such ways. (There are at least 10,000 known alkaloids, and alkaloids are only one of many classes of secondary compounds.) These chemically potent compounds tend to be beneficial or pleasant in small doses, but toxic in larger quantities. Yes, humans have used them as murderous poisons, too.

Optional sidebar on “fiction meets real life”: If you have watched the chemically-sophisticated series *Breaking Bad*, you may remember that the plots of several episodes hinged on a compound called ricin, which was extracted from castor beans and reputedly fatal in tiny doses. Ricin is the real thing. I grew up knowing this because my father nearly died from inhaling some that a sloppy colleague had left in a centrifuge in the pharmacology lab they shared during WWII. More recently, ricin was in the news in June 2013 when a pro-gun nut was sentenced to 18 years for mailing it to President Obama and Mayor Bloomberg. In the 1950s, my father grew castor beans in our backyard garden, I guess as some sort of memento mori of his own. My brother and I were given the strictest possible warning to stay away from them!

Unlike bench scientists in white lab coats, field biologists were also keenly aware that—despite the world’s greenness on a gross level—plant tissues faced an unremitting onslaught from specialized animals: many mammals, some birds and reptiles, and thousands and thousands of species of insects. Various lines of evidence implicated *defence against herbivores* as the main adaptive function of secondary compounds, and that has become the accepted interpretation.

Furthermore, although such chemicals may be “secondary” in the original sense of the term, their contribution to plant survival and fitness is hardly secondary. Individual plants that lack protective chemicals, either through artificial breeding to be palatable crops or through knockout genes, are highly susceptible to attack.

The coevolutionary arms race

You may sense an internal contradiction in the previous paragraphs. If these plant defences make life so hard for herbivores, why is herbivory found in so many animals, especially insects? The brief answer is that herbivory came first and diversification came second. In a brilliant and classic paper from 1964, Paul Ehrlich and Peter Raven showed that the evolutionary relationships between plants and the insects that eat them give clear evidence of a *coevolutionary* interaction that leads to specialization and speciation. You have learned about speciation in the evolutionary half of this course, but coevolution between antagonists gives the process more urgency and impetus.

A plant species is fed upon by an insect species. A mutant plant pops up that makes a toxic chemical that deters the herbivore. Because it is advantageous, the poisonous trait spreads in the plant population. However, a mutant insect turns up with the ability to detoxify the compound, and that trait spreads through the insect population. (We know that insects can overcome chemical defences quickly, because of their documented track records of speedily evolving resistance to synthetic pesticides like DDT.) Because the plants are now vulnerable to the insects that have coevolved to circumvent their defences, the cycle can repeat itself with a new plant compound. At any point, plant populations in different parts of the species range might evolve different compounds, and their local insect populations would track those compounds. That can lead to local specialization, and that can lead to different lineages differentiating enough to become different species. (Jerry Coyne is an authority on speciation. Review *Why Evolution is True* for details on the process.) After an insect species has developed resistance to a particular compound, it can evolve to specialize on that plant, sometimes even using the compound as a chemical cue to help find the plant, as a feeding stimulant, or as a cue for egg-laying. In many cases, such as monarch butterflies, a resistant insect can incorporate a poisonous compound into its own tissues to contribute to its own chemical defence against its predators.

Ehrlich and Raven deduced that this scenario must apply, and their synthesis inspired a huge amount of research that has supported it. Part of their evidence lay in *parallel adaptive radiations*. These are not limited to the butterflies that Ehrlich and Raven focused on. For example, beetles in the Family Bruchidae develop in the seeds of legumes, killing the seed in the process. There are many species of legumes; like castor beans loaded with ricin, many of them have toxic chemically-defended seeds, but also have a specialist bruchid that can develop in the seeds despite the potent chemistry. The diversification process evidently proceeded in parallel, with back-and-forth escalation of counter-adaptations. Bruchid diversity and legume diversity are linked.

Specialization is the rule rather than the exception for insects that eat leaves or otherwise harm plants. Interactions that are not so hostile, such as that between flowering plants and pollinating bees, are more generalized, another bit of evidence that speciation is augmented by antagonistic interactions.

More recently, modern molecular methods for inferring evolutionary trees have provided strong support for the idea that the evolutionary arms race of adaptation and counter-adaptation is indeed responsible for generating the astonishing diversity of insects and flowering plants. It is possible to locate branching points in the phylogeny of insects at which one branch became herbivorous while the other remained carnivorous. The herbivorous lineages have typically produced more species than their carnivorous sister lineages, although they necessarily have had the same amount of time in which to evolve.

This pattern of extreme specialization does not appear in large-bodied herbivores such as mammals, who experience plants very differently. Most insects, such as caterpillars whose mothers had to choose some particular plant on which to lay their egg, will spend their entire larval lives on a single plant individual. Whatever defences that plant has, the caterpillar must handle them or die, because it has no options. In contrast, a bison or a moose can move through its habitat and make choices. It can take many different species, sometimes getting several plant species in a single bite. Therefore, these large grazers can handle toxic chemicals by selecting a mixed diet in which they avoid the worst poisons and take milder toxins in low enough doses to avoid bad effects. They are diet generalists. They typically don't need—and usually don't have—highly specific detoxification pathways for dealing with certain secondary compounds.

Part 10. Synthesis and conclusions

I began this introduction to the ecological aspects of adaptation and diversity by stressing that the world is a heterogeneous place. Spatial structure—patchiness—is everywhere. So is the temporal equivalent of patchiness. We have seen how interactions have time lags, and how the frequency of disturbance events can be as important as their magnitude. I conclude my sequence of lectures on basic ecological concepts and processes with a case study that is designed to look back at several of those processes. I show that pocket gophers interact with lilies in a complex way that depends on the plant’s life history and the abiotic properties of the environment.

I hope that I’ve made it clear that patchiness is not only ubiquitous but that the spatial and temporal scales of patchiness are important. On the one hand, they affect interactions. For example, gophers have a clearly negative direct effect on individual adult lilies: they eat them and kill them. But at a different time in the plant’s life history, gophers probably benefit lily seedlings: by tilling the soil, they bury seeds that would otherwise desiccate and die. And, at a larger spatial scale, they almost certainly help lily populations persist by preventing succession that would otherwise turn open meadows into closed, shady forests. The overall interaction between gophers and lilies is not a black-and-white picture of heroes and villains. And this is not some rare peculiarity of gophers and lilies. There is every reason to assume that all interactions will reveal scale dependence if we are able to study them carefully enough.

We have also seen how metapopulation structure, with infrequent migration rather than complete homogenization, also changes the outcomes of interactions. Competitive or predatory relationships that would lead to extinction in homogenized populations can lead to persistent coexistence in weakly coupled metapopulations. In a somewhat analogous way, occasional disturbances that knock back succession can also foster coexistence and species diversity. (*Pop-culture mnemonic: Disturbance plus metapopulation structure means that natural communities are like James Bond’s preferred martinis: shaken but not stirred.*) Spatial patchiness and chance events in time are critical to the ecological outcomes. They also determine species characteristics through evolution. Species are adapted to the spatio-temporal graininess of their landscapes, not just to average or typical conditions. In lab-bench sciences, these effects are deliberately removed from consideration; indeed, organisms are plucked from their habitats entirely. There are valid reasons for doing that, as a part of pursuing biological knowledge. *But in-*

vestigations that never consider an organism's environment will never fully understand the organism.

In addition to trying to convince you that ecological contexts matter, I have also tried to give you some insight into the process of evaluating scientific ideas. I have probably been too hard on Frederic Clements, who definitely contributed innovative insights to science. I chose to hold him up to scrutiny because his career illustrates a dangerous tendency in the development of theoretical ideas: enthusiasm can easily slide into dogmatism. Enthusiasm-plus-dogmatism is a potent combination for recruiting fellow enthusiasts. It is hard enough to see the world objectively in any case. It is harder still when you are invested in a pet idea. In such a situation, you need to exert yourself to look for contradictory evidence. It is all too easy for bad theory to become accepted.

Clements had an additional weakness, although it was typical for his time. His grasp of evolution was limited. Indeed, when he did experiments that involved transplanting new species into gardens at different elevations in the Rocky Mountains, he reported that the new species had transmuted themselves into the closely related species that were previously at the site. He saw this as powerful evidence for the moulding power of the environment. But it couldn't be correct. Species don't transmute. Presumably, his transplants died and were replaced by local seedlings; he just lost track of which plants were which. A better understanding of evolutionary ideas might have made him more skeptical of his flawed ideas about community integration.

Parallels from medicine

For better or worse, the resolution of the Gleason-Clements debate hasn't exactly resonated throughout society; it hasn't touched all that many lives. Many of you are interested in medical careers, so let's look at some parallels. For an especially good example of bad theory in medicine, consider the ancient "doctrine of signatures." This idea, which goes back at least to the first century AD, held that plants whose parts looked like parts of the human body would provide medicines that were good for treating diseases of those parts. Therefore, the spring ephemeral wildflower *Hepatica*, whose trilobed leaves vaguely resemble the lobes of the human liver, must be the drug to use for liver ailments. The rationale for this theory was that a benevolent God would give humans signals about what drugs to use. To a post-Enlightenment mind, this rationale seems spectacularly weak—why wouldn't a benevolent God just prevent our liver ailments in the first place, for example?—but it held sway in Europe for centuries, and was most actively promulgated in the 1600s. Science finally banished this nonsense. Still, it had more staying

power than one would expect, given that it was an utterly useless notion. Why did the doctors of the day keep doing something that didn't work?

OK, the doctrine of signatures flourished at a time when some Europeans could be burned to death *just for asking* heretical questions about why God might do one thing instead of another. The whole point of something being a *doctrine* was that it should not be questioned. It was a bad time for scientific skepticism. But presumably we have left the Dark Ages behind, and medical practice is now informed by the bright light of science rather than untested propositions that don't work. If you think that, you might want to read *The Emperor of All Maladies*, a history of cancer by Siddhartha Mukherjee. He recounts the emergence of William Halsted's theory about surgical treatment of breast cancer. The underlying observation was this: surgical excision of a breast tumour occasionally brought about a cure, but usually the disease came back.

Halsted surmised that the unsuccessful operations must have left some bad tissue behind at the excision site, so he pioneered the "radical mastectomy" in which the whole breast was removed. Tumours still tended to recur, so he started removing underlying chest muscles and even ribs. A sort of competition developed among surgeons to see who could remove the most tissue without killing the patient, and desperate women sought out the surgeons who promised to mangle them most thoroughly. Both doctors and patients believed that *the chance of beating cancer increased with the severity of the operation*, but Mukherjee contends that data were available to show that this was untrue.

In fact, the reason that these cancers came back was usually *not* that undersized excisions left cancer behind. The disease returned because cells had metastasized throughout the body before the operation. Most of the women would have fared as well with a minor lumpectomy. These painful, disfiguring, sometimes lethal operations had no better success rate than the doctrine of signatures. We can assume that these surgeons were intelligent and compassionate people. Still, the medical establishment could not be bothered to act like skeptical scientists instead of complacent doctors and *submit their theory to the evidence that was available*. Like Clements, Halsted was an eminent authority. Like Clements's ideas, Halsted's radical mastectomy doctrine had a plausible rationale. That was enough—enough to inflict unnecessary anguish on a whole generation of afflicted women.

If that story makes you angry, I'd say that you have the potential to be a scientist. When people make mistakes without malice, we tend to say, "Well, they're only human." But the power of science is that it gives fallible humans a better chance to get things right. That power could have saved Halsted from doing a lot of harm. But it requires consciously setting aside ego and self-assurance, and giving precedence to humility, skepticism, objectivity, and evidence.

Metastatic cancer as an invasive species problem

Here's a newer Siddartha Mukherjee story that links medicine and ecology even more explicitly. As I have been adding my 2017 edits, I've come across a new piece of his in the 11 September 2017 issue of *The New Yorker*. He explains how traditional oncologists have been convinced that cancer is caused by cells that have developed some defect that makes them metastasize. But various lines of evidence suggest a different view: that metastasis happens because those cells find themselves in an *environment* that produces metastatic behaviour. His explicit analogy is to the phenomenon of invasive species, and in fact his article starts with the ecological paradox that Eurasian species like zebra mussels and quagga mussels are *not* invasive in their home ranges, but they become invasive (= "metastatic") when transplanted to the Great Lakes. The mussels have not changed, but their ecological milieu has.

Mukherjee tells a parallel transplant story as follows: "A patient—call him D.G.—is diagnosed with melanoma, and successfully treated with surgical resection. Years later, D.G., now deemed perfectly healthy, donates a kidney to a friend. The friend is prescribed routine immunosuppressants to prevent the rejection of the kidney. A few weeks later, however, the recipient begins to sprout hundreds of black pinpricks of melanoma in the kidney. The melanoma, bizarrely, has come from D.G.'s cells. The donated kidney has to be removed. Meanwhile, the donor—like some Dorian Gray of transplantation—remains uncannily healthy, with no sign of melanoma in his body." What's going on here? D.G.'s continued good health is not simply a consequence of his long-ago surgery. Instead, his immune system has been holding the disease at bay. The recipient's immune system lacks the requisite history with the disease, and in any case, it has been shattered by the immunosuppressants. Mukherjee develops an extended ecological analogy: what determines whether a plant grows—*the seed or the soil?* Cancer geneticists would search for explanations in the seed, but ecologists would tend to look to the soil. Consider Tansley's transplant experiments, or Gleason's assertion that plants will be found wherever conditions are right, unless prevented by dispersal. Mukherjee's conclusion is that doctors have to start thinking like ecologists:

“Once we think of diseases in terms of ecosystems, then, we’re obliged to ask why someone didn’t get sick. Yet ecologists are a frustrating lot, at least if you’re a doctor. Part of the seduction of cancer genetics is that it purports to explain the unity and diversity of cancer in one swoop. For ecologists, by contrast, everything is a relationship among a complex assemblage of factors...Going further and embracing an ecological model would cost us clarity. But over time it might gain us genuine comprehension.”

Well, “we need ecology” is a refreshingly humble and sensible statement to emerge from the medical establishment. I think that “genuine comprehension” is in fact the true goal of science, so Mukherjee’s argument resonates strongly with me. In fact, it is designated as a required reading for BIO120. A link to the article will be available on Blackboard, or you can find it here:

<https://www.newyorker.com/magazine/2017/09/11/cancers-invasion-equation>

Whether or not you accept the argument for ecological oncology, the article will provide you with impressive fodder for job interviews in medical labs!

How you need to be

When I was your age, it was the late 1960s—a turbulent time of war, protests, riots, and assassinations. Two admonitory bumper stickers were ubiquitous. One said “Never trust anyone over thirty.” Now that I’m an old geezer, I’m not sure how well that one holds up. I would prefer to tweak it a bit: “Never trust anyone who doesn’t offer good evidence.” The other was “Question Authority.” As an old geezer and a scientist, I can tell you for certain that that one still applies and always will. And turbulent times are back again.

Is there life after BIO120?

Looking back on my own university courses, I gained from most of them. A few of them made little impression; I emerged from the final exams feeling unchanged. And I remember two courses—Bacterial Genetics and something on the biochemistry of algae—that actually made me feel dumber. Whether the material was over my head or just poorly presented, I felt baffled and unengaged. I hope that BIO120 has not done that to you. I want it to be a course that makes you feel *smarter*.

If you are taking BIO120, you probably have some ambition for a career in the sciences. I consider that a worthy aspiration, and I hope you succeed. I also hope that you absorb and follow the lessons I’ve offered about the need for humility and skepticism in science, and the need to confront your hypotheses with objective evidence. Be ambitious, but be

a Gleason rather than a Clements. Learning to tell good science from bad is your responsibility. If you are on a professional track, my impression is that medical curricula don't include experimental design (no time! no time!), so you need to master that art as an undergrad. Take the hardest-nosed biology and statistics courses that you can find.

Science done right is a powerful and ethical way to harness the power of the human mind for establishing the truth. I think that we can use the word "noble" to describe science at its best, but science done wrong is just shabby careerism under another name. Like the humanities, science should also be a platform for experiencing the simple pleasure of knowing things. *Ecology done right* opens the widest and brightest window on our beautiful world.

Literature Cited

Allee WC, Emerson AE, Park O, Park T, Schmidt KP. 1949. Principles of Animal Ecology. Saunders, Philadelphia, USA.

Forseth I. 2010. Terrestrial Biomes. Nature Education Knowledge 3(10):11

Mukherjee S. 2017 Sept 11. Cancer's Invasion Equation. The New Yorker 93(27): 40.

