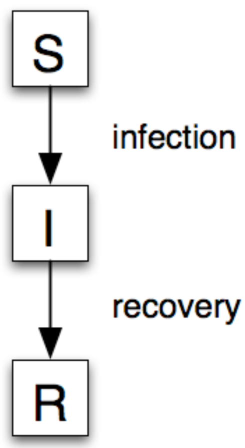
BMA 771: Introduction

An Example Model and Some Terminology: SIR model for the spread of an infectious disease.

This **compartmental** model describes the numbers of Susceptible, Infectious and Recovered individuals in some population.

Two representations of the model:

- A flowchart that shows how individuals move between classes. Arrows depict biological processes (infection and recovery) that lead to movement of individuals between classes.
  (We could add other arrows to show how these flows are dependent on the numbers of S, I and R.)
- A set of differential equations, which describes the rates at which the flows between classes occur.



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

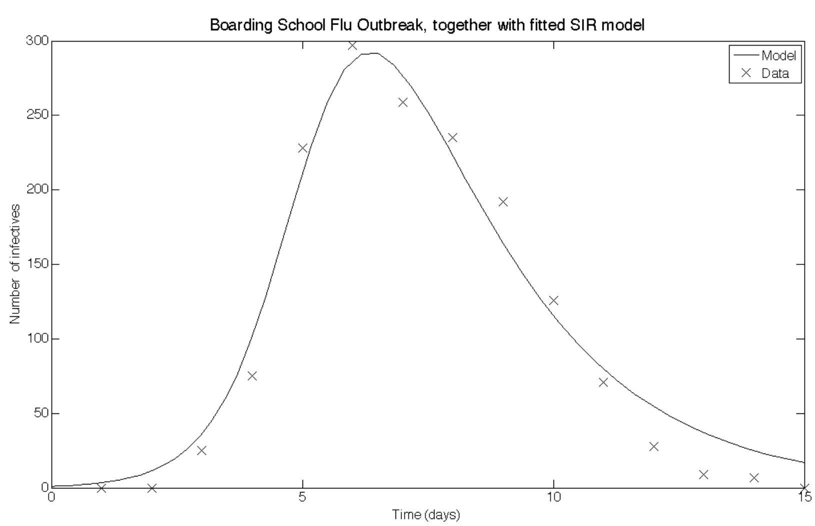
$$\frac{dR}{dt} = \gamma I$$

State variables: those quantities that specify the state of the system, i.e. what we need to know in order to say how the system will change over time (S, I and R).

The **dimension** of the model is the **number of state variables**.

Parameters tell us something about how the flows depend on the state variables (e.g. per capita recovery rate relates the number of Is to the rate at which I individuals move into R).

Typically, parameters will be estimated from data: ‘model fitting’, ‘parameter estimation’. Statistical input here. (How well can we estimate these parameters?)



Why Model?

1. **To make predictions**
   Perhaps the best-known examples are the models used to make weather forecasts.

2. **To do experiments** that might not be easy (or even possible) in reality
   'What if' questions:
   What would be the implications of removing a species from an ecosystem?
   Or introducing a species?

3. **To better understand a system**
   What biological processes are important?
   What is the role of this or that process?

   Modeling provides a framework within which we can make sense of data. (For instance, the large amount of data generated by molecular biologists, such as microarray data, has stimulated modeling of cellular and molecular systems: often called 'systems biology')

4. **To guide experimentation and/or data collection**
   What additional information do we need to know?
   Is the behavior of the model sensitive to particular parameters or processes? If so, we need to know these in more detail than if the model were not sensitive to them.

In order to make a model, we have to think in a very precise way since we are formulating problems within a mathematical framework.

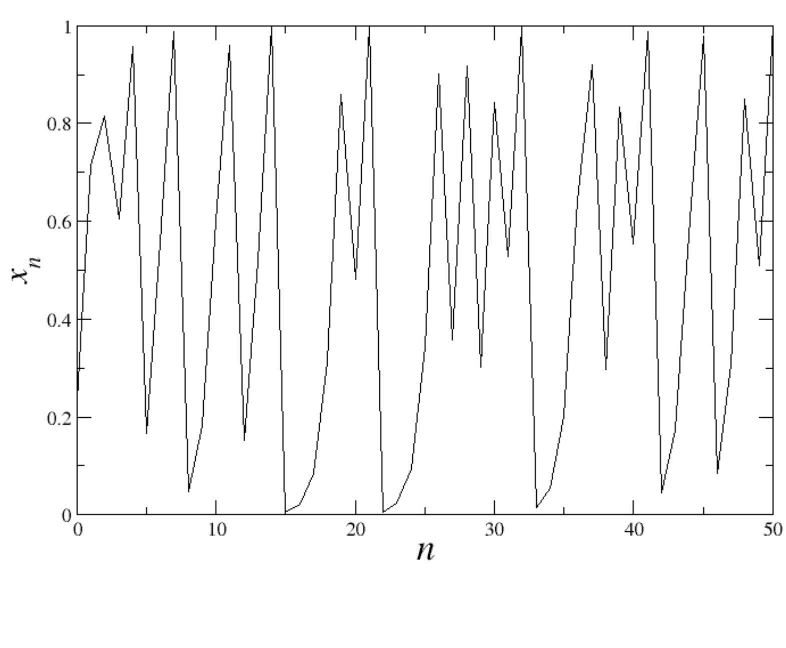
This can help us to pose questions in a well-defined way.
Can also help us to put ideas together (integration/synthesis)

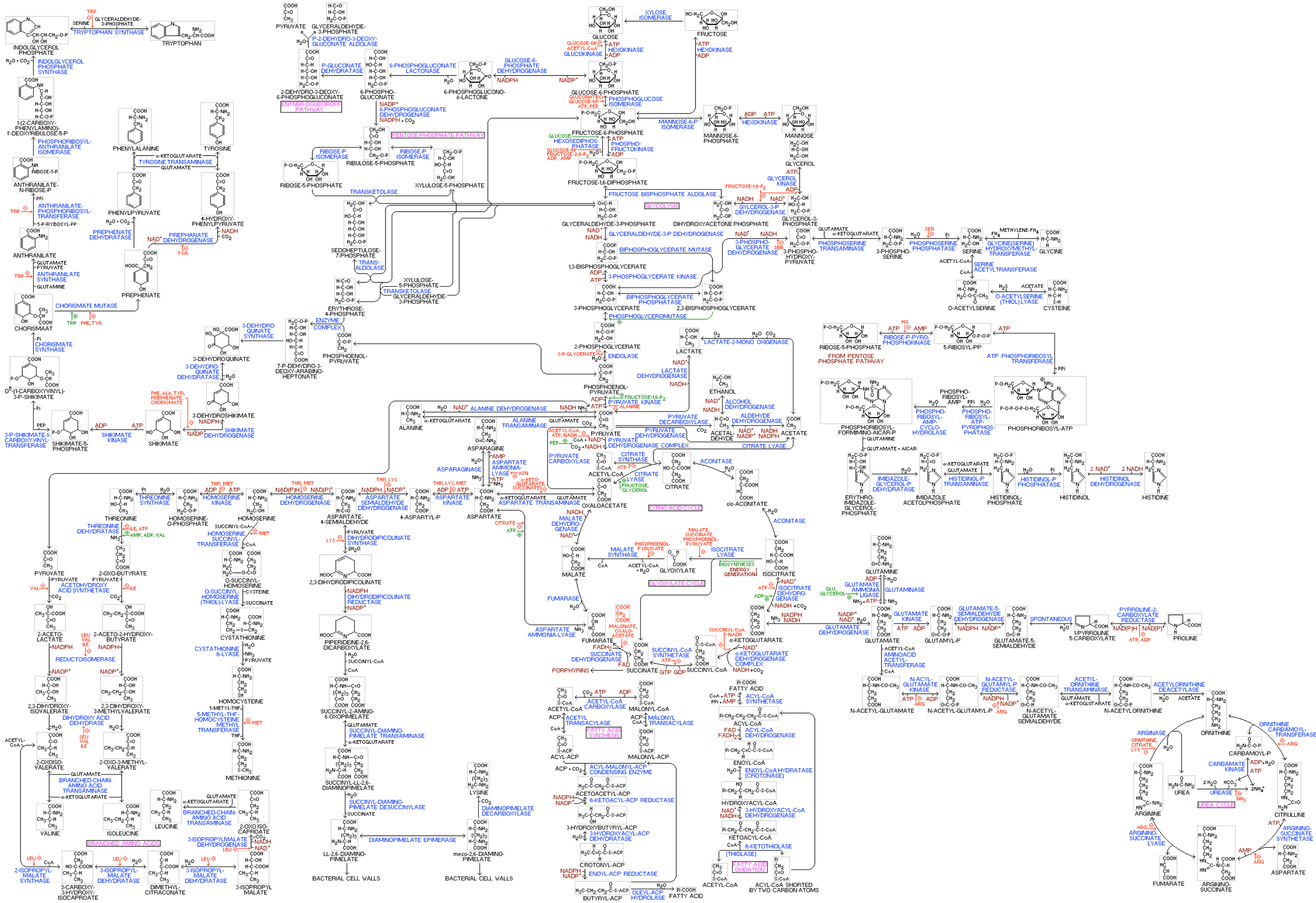
Important Question: Can we understand a complex system by studying its parts (e.g. cell metabolism figure overleaf)?
This **reductionist approach** dominates many areas of science.

In many settings, we see that **interactions—even between simple components— can lead to complex, often counter-intuitive, behavior.**
linear vs nonlinear systems

Having a simple model does not imply there will be simple behavior!

Classic example: logistic model (discrete time) $x_{n+1} = rx_n(1 - x_n)$ for $r = 4$.




ENTNER-DOUDOROFF PATHWAY
PENTOSE PHOSPHATE PATHWAY
GLYCOLYSIS
CITRIC ACID CYCLE
GLYOXYLATE CYCLE
FATTY ACID SYNTHESIS
FATTY ACID OXIDATION
UREA CYCLE
BRANCHED AMINO ACIDS
TRYPTOPHAN SYNTHASE
TRYPTOPHAN
PHENYLALANINE
TYROSINE
CHORISMATE MUTASE
PREPHENATE
SHIKIMATE
GLUCOSE
FRUCTOSE
GLUCOSE-6-PHOSPHATE
FRUCTOSE-6-PHOSPHATE
FRUCTOSE-1,6-DIPHOSPHATE
FRUCTOSE BISPHOSPHATE ALDOLASE
GLYCERALDEHYDE-3-PHOSPHATE
DIHYDROXYACETONE PHOSPHATE
3-PHOSPHOGLYCERATE
2-PHOSPHOGLYCERATE
PHOSPHOENOL-PYRUVATE
PYRUVATE
LACTATE
ETHANOL
ACETYL-CoA
CITRATE
cis-ACONITATE
ISOCITRATE
α-KETOGLUTARATE
SUCCINYL-CoA
SUCCINATE
FUMARATE
MALATE
OXALOACETATE
GLYOXYLATE
ASPARTATE
ASPARAGINE
ALANINE
THREONINE
HOMOSERINE
METHIONINE
LYSINE
VALINE
LEUCINE
ISOLEUCINE
SERINE
GLYCINE
CYSTEINE
HISTIDINE
GLUTAMATE
GLUTAMINE
PROLINE
ORNITHINE
CITRULLINE
ARGININE
UREA
FATTY ACID
BACTERIAL CELL WALLS
DIAMINOPIMELATE EPIMERASE
ATP PHOSPHORIBOSYL TRANSFERASE
PHOSPHORIBOSYL-ATP

Types of Models. There are many different ways to classify models, including:

Quantitative or qualitative

Do we care about numbers (e.g. prediction) and detailed behavior
Or more about general patterns of behavior (e.g. if I change X, does Y increase or decrease?)

Mechanistic or phenomenological (or statistical)

Model involves some description of biological processes
Or the terms of model need not represent real processes

[Example: weather forecasting. Can we predict tomorrow's weather in terms of past behavior, without resorting to modeling atmospheric processes?]

Continuous or discrete time

Time varies continuously (e.g. differential equation) or discretely (e.g. difference equation). Examples of latter often involve species that have non-overlapping generations, e.g. annual plant species.

Deterministic or stochastic

Does our model include random (stochastic) effects? Or do we imagine that the world runs like clockwork?

Population level or individual based models

Can we just model the numbers of individuals of different types, or do we need to keep track of each individual as a separate entity?

The type of model and level of detail we should use depends on what we want to do with the model, what questions we want to ask of it and how much we know about the system.

Some details may not be important when we ask some questions, others may be.

Another example: if we want to ask about (say) the spatial distribution of animals, our model needs to include some spatial component.

These issues are often discussed under heading of **Simplicity Versus Complexity**.

Simple models:

Are easier to understand, analyze, simulate, … more likely to give general insights
Have fewer parameters (unknowns to be estimated from data)
But are less likely to be accurate descriptions of reality

Complex models:

Are much more difficult to understand, analyze, simulate
More likely to be accurate descriptions of reality (provided that the assumptions of the model are correct)—but may be more specific to a given situation
Involve more parameters. Can be prone to overfitting (example).
Famous quote from Freeman Dyson:

> In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

Model selection is often guided by the principle of parsimony

We prefer the simplest model consistent with reality (Occam's Razor)

Modeling: an iterative process

Once you have made your model: compare with reality.

Are there discrepancies between predicted behavior and what is observed experimentally? If so, refine the model.

Modeling is an iterative process, with the model being refined at each step.

Notice that we often learn more when a model doesn't work: we know that we have missed something important.

A model can fit well, but for the wrong reasons… it might not capture the mechanisms that underlie the system.

Model sensitivity

How much does the outcome of the model depend on…

- Values of the parameters that appear in the model?
- Assumptions made in the formulation of the model? Structural sensitivity.

How much faith do we have in the answer that the model gives us?

Do our assumptions affect quantitative predictions? Or worse, do they affect qualitative predictions?

Message: Start simple and add in complexity, if needed

When presenting a model, it is important to lay out the assumptions that underlie the model. This will help you (and others) to criticize the model: what are its potential weaknesses?

Need to compare model to reality.