**Model Specification  
(determining the nature of model elements and the form of the rate and auxiliary variable equations)**

In this class, ***Model Specification*refers to determining which elements will be storages, flows, and auxiliary variables, and determining the form of the equations that specify the mathematical relationships between model elements**. The objective is to select equations that will approximate the actual relationships between the elements in the real system.

For example, if in the real system we believe that an increase in A causes a similar increase in B,

* then we would probably consider using a linear equation

Conversely, if we believe an increase in A causes a decrease in B,

* then we would probably consider using an inverse equation of some sort

Whereas storage variables are computed by integrating their rates of change, the values of auxiliary variables and rates are computed instantaneously at a given point in time.

It is the form of these "instantaneous" equations that we seek. Of principle interest are:

* different forms of relationships between two elements, y=f(x)
* and commonly used formulas involving more than two elements, including
  + summations (z=x+y+...)
  + ratios (z=x/y)
  + differences (z=x-y)
  + products (z=x\*y\*...).

Some useful forms of relationships between two elements include ([click here for sample graphs](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/CalibrationGrphs.htm?ou=676153) of eqns below):

* **positive, linear**:
  + y = a + b\*x
* **positive, but diminishing** (as x gets bigger, the change in y gets smaller):
  + y = a \* square root of b\*x
  + y = a \* ln (b\*x)
* **negative linear**:
  + y = a - b\*x
* **negative, but diminishing**:
  + y = a / x [inverse relationship]
  + y = a \* exp ( -b\*x )
* **positive, and amplified** (as x gets bigger, the change in y gets bigger yet):
  + y = a \* x^b
* **positive, but with definite limits**
  + y = a \* x / (1 + b\*x )
    - x and y are non-negative, b is often left at 1
    - y increases rapidly with x near the origin, but approaches the value a/b as x approaches infinity
    - shape is controlled by choice of a and b
    - could also raise x to a power c (in both numerator and denominator) to adjust shape further
  + y = a \* exp (bx) / (1+exp (bx) )
    - x goes from - infinity to infinity
    - y goes from - a to a
    - y is a/2 at x = 0
    - b is an optional scaling parameter
    - this equation is a bit more complicated, and is included because such formulae can be useful for characterizing qualitative variables
    - don't be concerned if it is not clear how you might use this; I've added this to the notes because it was used recently in a Ph.D. dissertation

Where do these equations come from?

* laws, pseudo laws, conjecture, guesswork, analogy, etc.
* if you are modeling a known system, they could come from measurements of the system or from theoretical knowledge about the relationships
* if you are modeling a system that does not exist and/or one that involves "soft" variables, the relationships will have to come from analogy about similar systems, from theory, or from experimentation with the model
  + in the latter case, part of the learning that comes from the model may come from changing the nature of element relationships to see the effect on model behavior
  + there are additional comments re "soft" or qualitative variables later in this section

**Read Sterman, Sec. 12-12.1.2, 12.1.6 - 12.2.1** (30 pgs.), **Sec. 13 - 13.2.5** (11 pgs.), **Sec. 13.3 - 13.4** (4 pgs.)  
Skim Sterman, Chaps. 14, 15, 16 (for awareness only)

**Generic Stock & Flow Templates and Infrastructures**

* Templates of the type described below can be very useful starting points when specifying models

[link to source doc](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/IST7.pdf?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153) (This document--optional material from Barry Richmond--has chapters specifically describing generic templates and structures.

Similiar to the concepts in Sterman but the terminology is different

A brief summary is provided below, to see if the material seems interesting to you...

**Generic Flow templates**

* compounding (self-reinforcing, growth, exponential)
* draining (decay, aging, half-life)
* producing (a resource stock is multiplied by a productivity factor to determine a production rate)
* co-flowing (the rates of two stocks are linked; coupled processes; tracking another activity or intervening effect; sales/revenue; code/bugs; conversion process)
* stock adjusting (a stock's value is compared to a goal, then divided by a TC to compute its rate)

**Generic Infrastructures**

* first order linear (compounding & draining together; stock with in rate and out rate each linked to stock value)
* s-curve (adds link from stock to the growth fraction or drain fraction)
  + I think there are better ways to model this
* overshoot & collapse (add Resource stock with its outflow linked to main stock; resource impacts growth fraction)
  + I don't like the graphical functions
* oscillation (two stocks, not linked by flow, impacting each other's rates; symmetric coupling)
* main chain (two or more stocks flowing from one to the next, e.g., a population chain of various cohort groups)

**Model Calibration  
(determining the proper values for the parameters in the model)**

**Parameters** include:

* constants modeled as auxiliary variables
* constants imbedded in equations
* the initial values of stocks

Where do parameter values come from?

* if you are modeling an existing system or a new one that incorporates well-known sub-systems, the values may come from physical laws and/or from previous experimentation with the real system.
  + consider the "thermostat model" where the relationship between temperature and units of heat is defined by equations from physics
* when you are dealing with soft variables, the values will often be defined completely at your discretion.
  + for example, if you are modeling anxiety, unless you have some experimental basis or measure of anxiety from which you want to adopt a scale and unit of measurement, you will have to define the range and meaning of anxiety values within the model.
  + this additional task is one of the reasons that modeling of soft variables is more difficult.

It is often necessary to "back into" (estimate) desired constants based on knowledge of system

* During initial runs, the numerical values may not be reasonable
  + for example, let's say you have created a model of an automotive cruise control system
  + it would not be reasonable for the model show that the car is traveling at 500 miles per hour, even if the feedback structure of the model correctly keeps the speed constant
  + if the model's feedback structure and the form of the equations specified are correct, then somewhere in the model there is a parameter that is not properly calibrated

Sometimes analysis of units can help with calibration:

* for example, if exercise is measured in sessions/week
* and we need to convert this into hours of exercise/week (in order to compute, say, energy burned)
* we know that the conversion constant has to be hours/week over session/week = hours/session
* which would lead us to approximately 1 or 2 hours as a starting point for the parameter to multiply "sessions/week" by to get "hours of exercise/week"

**Initializing in Steady State**

As discussed by Barry Richmond "An Introduction to Systems Thinking" 2001, sometimes it is useful to initially calibrate your model so it starts in steady state. The example provided might be helpful. (2 pages) These files open in MS Word so you can print them easily. You may need to close a blank window.

[IST Appendix Page 1](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/appendixpage1.doc?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

[IST Appendix Page 2](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/appendixpage2.doc?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

**Discussion of Delays**

**Read Sterman Sec. 11.1-11.2.1** (4 pgs.) **11.2.3, 11.2.4** (2 pgs.) and **11.3** (8 pgs.)  and Skim rest of 11.2, but do not worry if some of this is not totally clear.

Review: Material Delays are physical flows and must always contain a storage

An example of an Information Delays could be the \*\*dawning\*\* awareness or recognition that something has changed

see Figures 11-10 and 11-11 in Sterman

Delys can be viewed as a smoothing function such as a Moving Average (MAV)

which might be weighted, often exponentially (called exponential smoothing)as the "order" of the delay increases, it approaches a discrete or "pipeline" delay see Sterman Figure 11-14VENSIM supports both smooth delays and discrete delays (DELAY built-in)

**How do we know when to utilize a feedback structure to create the desired behavior versus creating an equation that directly generates the desired relationship?**

* As we are using the term in this class, specification deals more with the latter (determining a direct equation)
  + but, it is not a critical distinction
  + in the end we have to formulate the entire model
  + some aspects will be implemented through storages and feedback structures
  + other aspects will be implemented through the specification of direct equations
  + we simply focused first on feedback structures and now on the specification of equations

To answer the question above for a specific case example, let's say that you are building a model that includes anxiety and debt, and you believe that anxiety is determined by debt level.

* Should you model anxiety as a storage and have debt influence the rate of increase of anxiety?
  + which would be an example of utilizing a feedback structure
* Or should you model anxiety as a function of debt (either proportionally, or diminishing, or amplified)?
  + which would be an example of specifying an equation

If you decide that anxiety should subside over time of its own accord even when debt is not changing:

* then, you might lean towards using a *FB structure*
* so that you can use a "drain" rate to implement the process of anxiety subsiding

However, if you decide that the anxiety level stays constant when debt stays constant:

* then, instead of anxiety being a storage whose values are determined through FB structures
* anxiety would be a modeled as an auxiliary variable containing a mathematical formula in terms of the debt level.

**Specification Example: Rat Starvation Rate as a Function of food supply** (adapted from a student paper)

* some laboratory data was available: feeding rate vs. deaths per day

|  |  |
| --- | --- |
| Feeding Rate | Deaths/Day |
| 10 | 20 |
| 50 | 5 |
| 100 | 3 |
| 500 | 2 |

* Rat starvation appears to be inversely proportional to food, but not linear
* How might one develop an equation that fits this data?
  + e.g., death fraction = K\* pop/food as an initial thought ??
* The modeler decided to use a table to try and "discover" a workable equation
  + played with different possible formulas (basically trial and error)
  + robustness goal: should work for all values for food
  + is/was quite a challenge
* Could have used a graphical function, but they tend not to extrapolate well

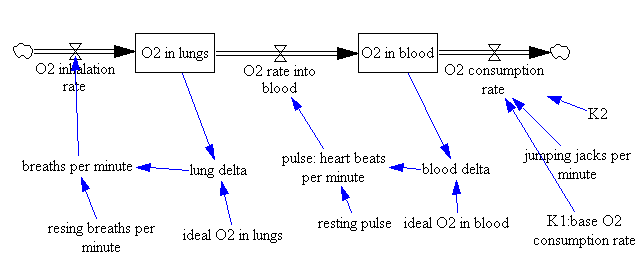
Rat population (P) is 50:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | what we want to compute | initial start at computing it | too big; scale by .1 | gets too small; take sq. root | rescale; mult. by .5 |
| Food (F) | Deaths (D) | deathfract (D/P) | P / F | .1 \* P / F | root (.1 \* P/F) | .5\* root (.1 \* P / F) |
| 10 | 20 | .4 | 5 | .5 | .7 | .35 |
| 50 | 5 | .1 | 1 | .1 | .3 | .15 |
| 100 | 3 | .06 | .5 | .05 | .23 | .12 |
| 500 | 2 | .04 | .1 | .01 | .1 | .05 |

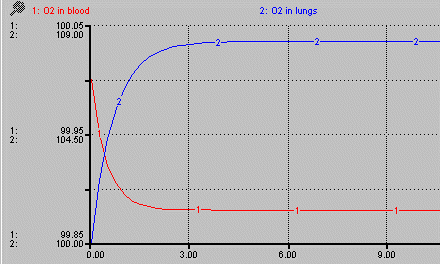
* Is this close enough??
  + compare the third column (the goal) with the final column (the candidate formula)
  + Actually, it's not all that good, but it does illustrate the thought process
* Result: rat death fraction = .5 \* (.1\*P/F)^.5)

**Example Specification & Calibration Model: Oxygen Flow**

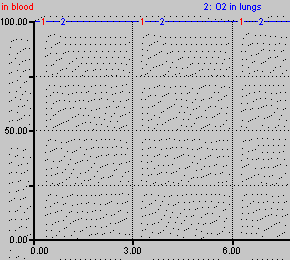
* Mainchain with O2 moving into lungs, from lungs to blood, from blood via exercise
* The first rate is proportional to breaths per minute
* The second rate is proportional to heartbeats per minute
* The third rate = K1 + K2\* exercise intensity (jumping jacks per minute)
* Pulse (heart beats per minute) = resting pulse + factor that increases as O2 in blood decreases
* Breaths per minute = resting breaths per minute + a factor that increases as O2 in lungs decreases
* Here is an initial flow diagram (before any equations are entered)
  + note: I added "lung delta" and "blood delta" to represent the difference between ideal O2 in the lungs and blood vs. actual O2.
  + these delta variables then influence breaths per minute and pulse (heart beats per minute)



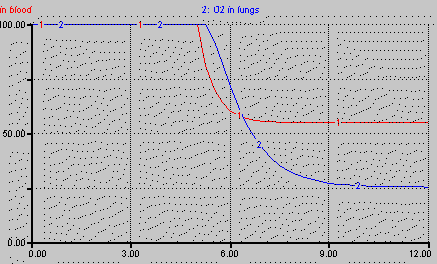
* Let's assume that at rest, O2 is evenly dist. between lungs and blood (we could find out through some research...).
  + Let's make this amount 100 arbitrary units (again, with some research, we could convert this to real units).
  + We'll set initial values and the ideal amounts for both lungs and blood to 100.
* Oops, in the diagram above, I forgot "resting breaths per minute"
  + I added it
  + and set the value to 15
  + I also set resting pulse to 60
  + I also forgot K2, so I added it as well
    - I don't know its value yet (we'll figure it out shortly)
* Since one can only live a few minutes without breathing, let's set K1 to 50 (half the amount of O2 in either storage)
* If resting pulse of 60 keeps us properly supplied with O2, then O2 rate into blood, must be pulse \* 50/60 = .83 \* pulse
* Now, let's assume that the max value for jjpm is 60, and, that at this rate the heart would eventually speed up to 150 bpm
  + since 150 is 2.5 times the resting pulse, O2 consumption at this point must be 2.5 \* 50 = 125
  + thus, K2 must be 75/60 = 1.25 in order to convert 60 jjpm into 75 O2 units per minute
* The lung delta and blood delta do not require calibration since they are simple subtraction (ideal minus actual)
* O2 inhalation rate must be 50 when breaths per minute = resting bmp of 15
  + so the coefficient in this flow equation would be 4 [this is not true, but I didn't catch my arithmetic error until later...]
* jjpm is an input function, and could be a constant, a step function, or whatever
* This leaves pulse (heartbeats per minute) and breaths per minute to calibrate
  + the key here is the time constant--how long does it take to react to changes
  + this is harder to pin down. As a starting point, let's imagine that 02 in blood has dropped to 80 from 100, so blood delta is 20
  + how much will pulse increase? O2 could be going out at 125 vs. an inrate of 50, so it might only take about 20 seconds for O2 to drop to 80
  + i.e., 20 seconds after beginning hard exercise, one's pulse could easily be up to 100
  + 100 = 60 + 40, or 60 + blood delta (20) times 2
  + thus, the formula for pulse = resting pulse + 2 \* blood delta.
* By a similar thought process, when lung delta is at 20, it is likely that breaths would have increased from 15 bpm to 20-25 bpm
  + thus, bmp is likely to = [approximately] resting bpm + .3 \* lung delta
* We will set jjpm to 0 initially and see how the model runs.



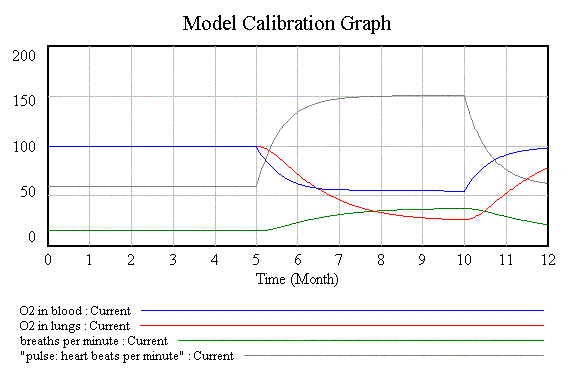
* As one can see, O2 in blood is almost exactly calibrated for steady state; O2 in lungs is a bit off. Why?
* It is because I made an error earlier when I divided the O2 inhalation rate of 50 by 15 bpm and got 4; the correct value for the coefficient in the O2 inhalation rate is 3.3
  + I'll make this change
* Now the storages stay at 100 (note that I have now set the plot scales to be 0 to 100):



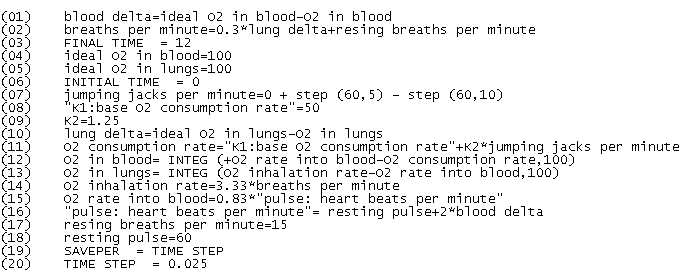
* Now I'll set jjpm = 0 + STEP(60,5), meaning jjpm goes instantly to 60 at time 5.
* Here's what happens:



* It appears that the increase in heart rate and breathing are not enough to replenish the O2. I will try doubling the parameters in the formulae for pulse and breaths.
* I get the same shape, except that the values stay close to 100. Maybe a linear correction factor is not sufficient??
* Wait, maybe I am plotting the wrong thing. O2 would decrease until the exercise stops
* Instead of continuous exercise, let's stop the pulse after 10 time units: 0 + STEP(60,5) - STEP(60,10) and change the time from 10 to 12.
  + I will also put the coefficients back to the original values reasoned out earlier



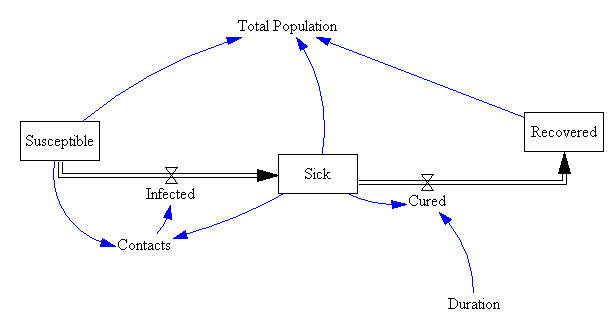
* Bingo! As a first pass, the behavior appears to be OK after all
* Here are the equations at this point:

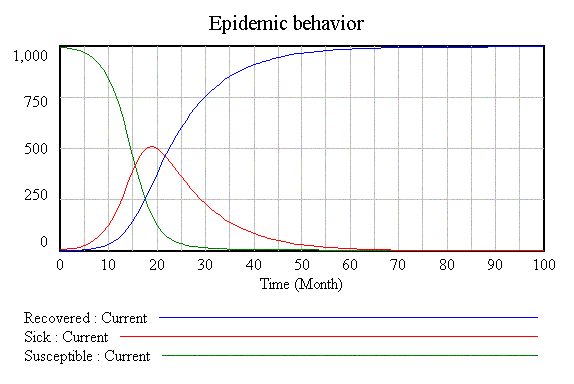


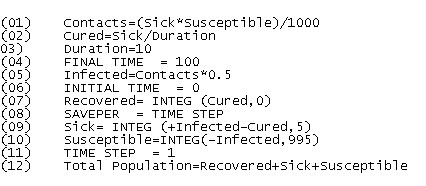
* It must be said that even though the behavior resembles my assumed reference behavior, the model is only a very crude approximation to reality; much more testing would be needed (and, most likely, significant additional model refinement) in order to be able to draw reliable inferences.
* The model is, never-the-less, a credible starting point; [link to the model as it now stands.](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/bloodflo.mdl?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

**Sample specification & calibration Model: Epidemic** [link to epidemic model file](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/epidemic.mdl?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

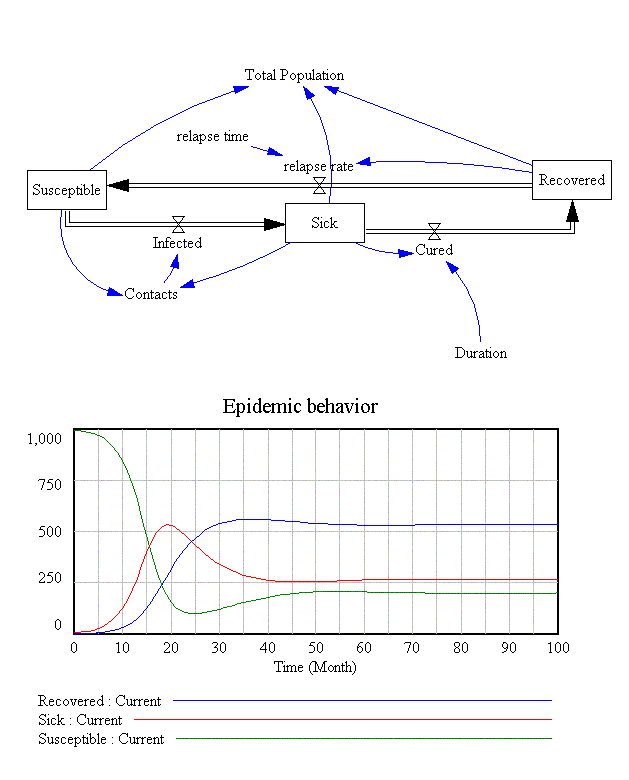
* mainchain, with susceptible, sick, recovered
  + start with 1000 susc, 10 sick, and 0 recov.
* the key rate (equation) is the infection rate
  + it is related to how many susceptibles there are and their coming into contact with infectious people
  + assume sick means infectious
  + first pass, make infection rate proportional to contacts
    - the proportionality constant depends on how infectious the disease is
    - for our model, we will use .5 times contacts
  + Contacts is proportional to susc. \* sick
    - this product is large when both are large
    - small when either is low
    - initially the product is 10\*1000=10000
    - a plausible formula for contacts might be the product over 1000, which yields 10 contacts per day initially
* recovery rate is simply sick/avg. duration of disease
  + assume avg. duration is a constant, say 10 days
* The behavior is S-shaped (downwards) for susc, S-shaped upwards for recov., and "bell-shaped" for sick





  
Adding a rate from recov. back to susc. with a large time constant

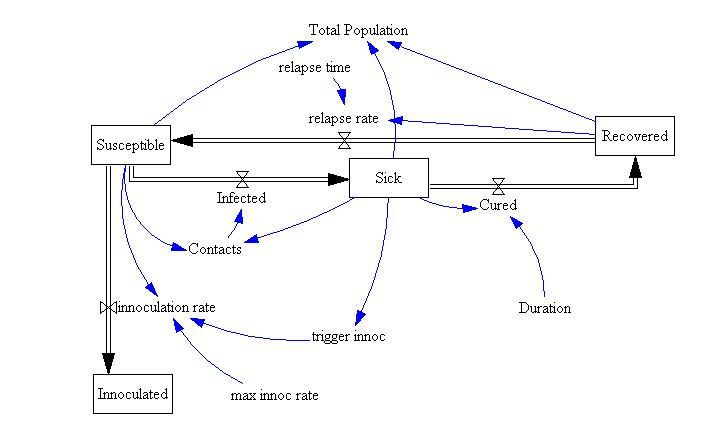
* to simulate people becoming re-susceptible

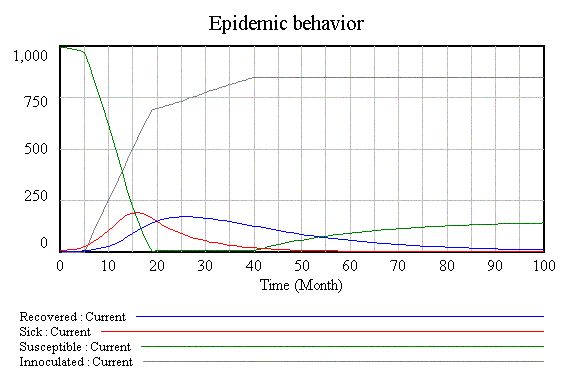


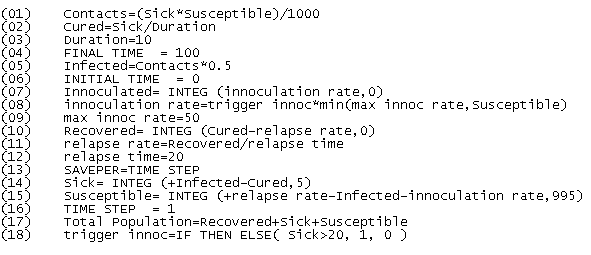
* now the epidemic never ends

Adding an inoculation program that triggers on sick or infection rate (recognition value)

* an interesting challenge would be to make this function asymmetric
  + don't shut inoculations off until well below the recognition value
* inoculation could be modeled as a flow from susceptible to a new stock called "inoculated"
  + assumes inoculated people are safe from the disease

[](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/EpidemicInnoc1.gif?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

[](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/EpidemicInnoc2.gif?_&d2lSessionVal=Q1KSCV9WJxlfvzb6ojdQHWKYp&ou=676153)

[](https://d2l.pdx.edu/content/enforced/676153-OFFERING_SYSC-514-001_201801/EpidemicInnoc3?ou=676153)

* As can be seen, the epidemic is avoided
* Also, due to the re-susceptibility, everyone ends up either inoculated or susceptible

Note that the trigger here is symmetric (shuts off at the same point as it goes on)